The Dissertation Committee for Amanda Barbara Chadee Certifies that this is the approved version of the following dissertation:

The roles of CYT-18 in folding, misfolding and structural specificity of the *Tetrahymena* group I ribozyme

Committee:

Rick Russell (Supervisor)

Karen Browning

Alan Lambowitz

Jon Robertus

Yuhui Whitney Yin

The roles of CYT-18 in folding, misfolding and structural specificity of the *Tetrahymena* group I ribozyme

by

Amanda Barbara Chadee, B.S

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2009

Dedication

To my parents, Ramnarine and Amryll Chadee

Acknowledgements

I would like to express my gratitude to all the people who have supported me throughout my graduate career. First, I would like to thank my advisor, Rick Russell for being instrumental in guiding me as a graduate student by fostering my analytical skills in planning experiments and for setting an excellent example on how to think critically about experimental results. I would also like to thank Pilar Tijerina for assisting me in protein purification, and Paul Paukstelis and Georg Mohr for always being available to answer all my CYT-18 related questions. To my committee members, thank you for all the insightful discussions and suggestions to improve my dissertation project.

To past members of the Russell lab, Travis Johnson, Jacob Grohman and Hari Bhaskaran, thank you for your warm welcome into the lab and for showing me how to perform proper lab duties and for introducing me to some great music. To the present lab members, Brain Cannon, Inga Jarmoskaite, David Mitchell, Cynthia Pan, Jeff Potratz and Yaqi Wan, you were all wonderful to work with, and I enjoyed all the wonderful stories and exciting scientific discussions.

I would also like to acknowledge my family for all their love, support and assistance during my PhD program. Mom and Dad, I appreciate all your hard work and the sacrifices you both made in order to provide me with the best educational opportunities. I am also thankful to my husband Siyavash. Your words of encouragement have always given me hope when things seemed down and your support continues to motivate me to be the best I can be.

The roles of CYT-18 protein in folding, misfolding and structural specificity of the *Tetrahymena* group I ribozyme

Publication No._____

Amanda Barbara Chadee, Ph.D.

The University of Texas at Austin, 2009

Supervisor: Rick Russell

Group I introns are structured RNAs that have been used extensively as model systems for RNA folding because they are experimentally tractable, yet complex enough to have folding challenges associated with larger RNAs. The *Tetrahymena* group I intron consists of a set of conserved core helices and a set of peripheral elements. Peripheral elements surround the core helices and form long range tertiary contacts between each other and to the core. Interestingly, a long-lived misfolded state is populated that has the same long range tertiary contacts as the native state but differs locally within the core.

Our lab showed that the intact periphery is necessary to specify the correct core structure, as mutating tertiary contacts or removing the P5abc peripheral element dramatically destabilized the native ribozyme relative to the misfolded form. However, we also showed that the thermodynamic benefit peripheral structure provided is accompanied by kinetic liability in folding, apparently because native tertiary contacts

formed by peripheral elements around the misfolded core must come apart to allow refolding of the misfolded RNA to the native state.

In addition to peripheral elements, proteins also play a role in stabilizing the native structures of many group I introns. The CYT-18 protein, which occupies the same binding site as P5abc, stabilizes the functional structures of certain group I introns by using a set of insertions that are absent in other related bacterial and mitochondrial aminoacyl tRNA synthetases. Using the P5abc deletion variant of the *Tetrahymena* ribozyme ($E^{\Delta P5abc}$), I sought to further define CYT-18 roles in RNA folding by probing its thermodynamic and kinetic effects on the native state formation relative to the misfolded state. I demonstrated that CYT-18, like P5abc, provided thermodynamic stability to the native state. However, unlike P5abc, CYT-18 had no apparent effect on the refolding kinetics, suggesting that a protein co-factor can stabilize the functional structure without acquiring the associated costs in RNA folding kinetics. Furthermore, I found that the mechanism of CYT-18 action appears to be distinct from P5abc. Disruption of the long-range contact P14, which is formed between P5c and L2 and is part of the network of peripheral contacts, dramatically weakened P5abc binding to the native ribozyme core by ~10^8 fold. Interestingly, CYT-18 maintained specific and tight binding to these mutants, which suggests that CYT-18 does not rely on a circular network of contacts to specifically stabilize the native state. Instead, the specificity may arise from a more direct and intimate contact of CYT-18 with the ribozyme core. This study gives insight into an evolutionary advantage of protein co-factors in RNA folding; proteins may offer thermodynamic assistance without inhibiting folding kinetics.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Alt P3- Alternatively paired region 3
ATP-Adenosine triphoshate
CMV- cytomegalovirus
Crm1- chromosome region maintenance 1
CTE- constitutive transport element
CYT- refers to cytochrome function in mitochondria.
DTT-Dithiothreitol
DMRIE-C- 1,2-dimyristyloxypropyl-3-dimethyl-hydroxy ethyl ammonium bromide-cholesterol
EIAV- equine infectious anemia virus
ER- endoplasmic reticulum
FCS- fetal calf serum
$\mathbf{E^{\Delta P5abc}}$ – P5abc deletion variant of the *Tetrahymena* ribozyme
EDTA-Ethylenediaminetetraacetic acid
GFP- green fluorescent protein
HERV-K- human endogenous retrovirus type K
HIV- human immunodeficiency virus
HTLV- human T-cell leukemia virus
IPTG- Isopropyl β-D-1-tigalactopyranoside
LB -media- Luria- Bertani media (1 % tryptone, 0.5% yeast extract, 0.5% NaCl)
LTR- long terminal repeat
MMTV- mouse mammary tumor virus
MOPS-3-(N-morpholino) propanesulfonic acid
mtLSU- mitochondrial large subunit
NES- nuclear export sequence
PAGE- polyacrylamide gel electrophoresis
P5abc-perpheral domain consisting of paired region (P).P5a, P5b and P5c
PEI- Polyethyleneimine
RcRE- Rec-responsive element
Rem- regulator of export/expression of MMTV mRNA
RmRE- Rem-responsive element
RRE- Rev-responsive element
RxRE- Rex-responsive element
SA- splice acceptor
SAFA- Semi-Automated Footprint Analysis
Sag-superantigen
SD- splice donor
SV40- simian virus 40
TE buffer- 10 mM Tris, 1 mM EDTA
TTE buffer- 70 mM Tris, 20 mM Taurine, 0.4 mM Na-EDTA

Chapter 1: General overview of RNA folding, misfolding and the influence of RNA binding proteins in stabilizing the native structure

1.1 FOLDING OF STRUCTURED RNAS

The biological functions of structured RNAs are widespread and diverse in the cell. RNAs can function as key regulators in gene regulation [1-4] and play structural and catalytic roles in pre-mRNA splicing [5] and translation [3, 6]. The importance of these RNAs in vital cellular processes warrants the study of the biochemical properties that govern their correct folding. For any structured RNA to be functional, it must fold into a specific three dimensional structure, which can be as simple as a riboswitch or as complex as the ribosome.

Whatever the spectrum of structural complexity, each RNA molecule has intrinsic properties that result in two main folding challenges: 1) to specify a single functional structure at equilibrium (thermodynamic challenge) and 2) to fold on a time scale that allows function (kinetic challenge) [7]. The thermodynamic challenge proves to be a problem at the most basic level of RNA structure, its side chains. The side chains are comprised of four bases that are grouped either as purines (adenine and guanine) or pyrimidines (uracil and cytosine). The limited diversity of these side chains gives RNA a relatively “low information content”, which limits its ability to specify the native state. Furthermore, the molecular information in its linear sequence decreases, because its base pairing side faces inwards from the sugar phosphate backbone, causing RNA helices to be very similar in shape, and thus making it even more difficult to specify a native state. The kinetic challenge for RNAs stem from the innate stability of RNA duplexes. These

duplexes can persist on a time scale from hours to days. This is a problem considering the strong propensity of RNAs to form alternative structures, such that the presence of non-native contacts can be trapped in a particular conformation. These kinetically trapped intermediates can limit the overall folding to the native state because these intermediates must at least partially unfold and then refold again before adopting the correct structure.

A large amount of the information we know today on RNA folding has come from *in vitro* studies of ribozymes by small angle X-ray scattering (SAXS) [8], time resolved hydroxyl radical footprinting [9, 10], single molecule fluorescence [11], selected 2'OH acylation primer extension (SHAPE) [12], computational analysis [13, 14] and ribozyme activity assays [15]. Though folding of an RNA molecule *in vitro* does not exactly model RNA folding *in vivo* (see Section 1.3), insights gained from *in vitro* studies undoubtedly shed light on how the information in a linear sequence of an RNA molecule governs its folding.

General steps involved in RNA folding include rapid formation of secondary structure (early folding) followed by consolidation of global structure through the formation of tertiary contacts (late folding) [16, 17]. Early folding events are generally several of orders of magnitude faster than late events, which can occur on the time scale of minutes. Folding is initiated through the electrostatic shielding of metal ions by neutralizing the negatively charged RNA backbone [18]. Results from SAXS and hydroxyl radical footprinting together demonstrated that the early folding events of the *Tetrahymena* ribozyme involved electrostatic screening and rapid collapse of the RNA [8, 19]

After a rapid collapse of the RNA, the bulk of the late folding events involve transitioning to the native state through rearrangement of one or more kinetically trapped intermediates. These kinetically trapped intermediates may have both native and non-native contacts that must be disrupted to allow further folding. These rearrangements may be fast, allowing RNAs to fold to the native state very quickly, or these rearrangements may kinetically trapped as evident from urea dependence studies (Figure 1.1) [20].

For many RNAs, the folding landscape traversed is intrinsic to its overall sequence. However; there is one additional feature which influences RNA folding, that is its interactions with proteins (see Section 1.6). Assembly with proteins impact the overall folding landscape for an RNA molecule by promoting the formation of the native structure [21]. Not only are proteins important during folding, but they play an integral part in structural stability, splicing and catalysis [22], since some proteins are retained in the final complex. For example, group I intron RNAs form tight complexes with proteins such as CYT-18 and CBP2 which function by promoting the formation of the native structure by acting as a scaffold for further assembly to the native structures or by "capturing" a collapsed native form of the RNA for activity [23].

1.2 RNA MISFOLDING

The discovery that slower transitions involving structural rearrangements can limit the overall folding to the native state led to further interest in understanding features of misfolding in certain RNAs. RNAs can misfold at the most basic level of structure formation, base pairing. The first evidence for the occurrence of alternative base pairing in RNAs was in studies of tRNA [24-27]. Gartand et al. [26] first demonstrated that

tryptophanyl tRNA could exist in two forms, only one of which could be charged with the correct amino acid. In addition, extensive probing of *E.coli* tRNA (Glu) in the absence of Mg^{2+} by diethylpyrocarbonate (DEPC), dimethyl sulfate (DMS) and RNase V1, revealed a hairpin structure with two internal loops. These features were different from the canonical clover leaf shape of tRNAs (Figure 1.2 A, B) [28], thus demonstrating that the misfolded RNA adopts a distinct fold from the native RNA.

The presence of alternative forms of RNAs extends beyond initial studies with tRNA. Extensive work has been done on studying the folding and misfolding of RNase P [29]. Bacterial RNase P functions in tRNA processing by cleaving pre-tRNA to generate their mature 5' ends [30]. Evidence for misfolding came from catalytic activity, circular dichroism and absorbance spectroscopy experiments, which indicated that this RNA required minutes to complete its folding despite rapid formation of at least one intermediate. Furthermore, folding was found to be accelerated by urea, a clear indication of a kinetically trapped species [31-33].

RNAs may also potentially misfold at the level of tertiary structure contacts. Since this step involves the ordering of secondary elements into conformations that are energetically stable, non-native contacts can feasibly form in the context of native structure [34]. In addition, there is evidence that another source of misfolding arises from differences in the spatial arrangement of helical elements such that non-native contacts are further reinforced by native tertiary contacts [32, 35].

While examples of RNA misfolding presented above result in structures that are non-functional, it is important to consider that alternative folds can have functional roles in the cell. For example, riboswitches located at the 5' end of mRNA can regulate gene

expression through structural transitions upon binding small metabolites. For example, the lysine riboswitch adopts a distinct structure in the absence or presence of lysine [36]. With lysine bound, a conformational change is induced in the "switching sequence", resulting in the formation of an anti-terminator stem to halt transcription. In the absence of lysine the "switching sequence" is sequestered and gene expression is turned on. Both folds are distinct, functional and are equally responsible for proper regulation of lysine. In addition, the Bartel group used rigorous mutational analysis and RNA selection to transform the HDV ribozyme and a synthetic RNA ligase ribozyme into molecules with 100% sequence identity. Remarkably, this RNA sequence resulted in two distinct folds that were equivalent in energetics and retained each distinct catalytic activity for site specific cleavage (HDV) and self-ligation (RNA ligase) [37].

1.3 RNA FOLDING IN VIVO

Though the study of RNA folding and misfolding in an *in vitro* system can give great insight into the principles of RNA folding, it is important to note that the intricacies of how RNAs fold in the cell may differ. RNA folding *in vivo* occurs co-transcriptionally and is probably highly influenced by the speed of RNA polymerase, site-specific pausing and interactions with proteins [38]. The vastly different external environment between studying RNA folding *in vitro* vs. *in vivo* called into question whether kinetically trapped intermediates or misfolded intermediates existed in the cell. Evidence for the existence of misfolding *in vivo* comes from the ubiquitous presence of RNA chaperones and DExD/H box proteins that function to unwind short RNA duplexes, suggesting they may also act

similarly on RNA structure *in vivo* [34, 39]. However, the extent to which these misfolded intermediates exist are still unknown.

Studies of co-transcriptional folding have been largely carried out by the Sosnick and Pan research groups [40]. Through a set of intricate experiments, they compared the co-transcriptional folding and *in vitro* folding of RNase P, which has two domains, a catalytic domain (C) and specificity domain (S) [41]. While they found that both folding processes formed kinetically trapped intermediates, the nature of kinetic traps were different. The kinetic trap in the Mg^{2+}-induced folding involved regions of both the C and S domain, while co-transcriptional folding allowed the C domain to fold before the S domain, though at a slower rate [42]. Interestingly, the addition of the elongation factor NusA accelerated folding of both domains by increasing the duration of transcriptional pausing to allow folding alterations to take place. Further work to elucidate the mechanisms of pausing and folding in RNase P demonstrated that pausing causes either local rearrangement of structure or may serve to sequester sequences until the correct base pairing partners arise, thereby preventing the formation of misfolded structures [32, 43].

1.4 GROUP I INTRONS: A MODEL SYSTEM FOR STUDYING RNA FOLDING

Coding sequences (exons) within genes are punctuated by non-coding regions (introns). Group I introns are widely distributed in lower eukaryotes and higher plants where they interrupt mitochondrial or chloroplast genomes [44]. Intron sequences, though initially transcribed to the precursor RNA, are eventually removed from the RNA by splicing out of the transcript. Indeed, one of the most remarkable discoveries was the

observation that the intron within the 26S rRNA from *Tetrahymena* catalyzed its own removal from the precursor RNA, a process referred to as self-splicing [44].

Group I intron self-splicing involves two chemical steps. First, attack of the 3' hydroxyl of an exogenous guanosine at the 5' splice site. This breaks the covalent attachment of the intron to the exon and the guanosine attaches itself to the 5' end of the intron. Second, the 5' exon attacks the 3' splice site, resulting in a free intron and ligated exons [45, 46] (Figure 1.3). Although many introns self splice *in vitro* under a variety of conditions, splicing of many introns *in vivo* is assisted by protein factors (Section 1.6).

In 1983 Cech et al. published a secondary structure model of the *Tetrahymena* group I intron from comparative sequence analysis and experimental constraints derived from single stranded and double stranded specific nucleases [47, 48]. Over time, features of this model were tested and confirmed via mutagenesis and phylogenetic comparisons. In groundbreaking work, Latham and Cech also defined the "inside" and "outside" of the RNA through solvent accessibility of hydroxyl radicals generated by the Fenton reaction [49], a technique that continues to be a powerful structural probe for RNAs. With the abundance of biochemical data generated, Lehnert *et. al* in 1996 constructed a three-dimensional model for the entire intron [14].

A key feature in studying RNA folding came from the transformation of the self-splicing intron into a self cleaving ribozyme by removal of the exon sequence (Figure 1.4 A) [50-52]. The *Tetrahymena* ribozyme cleaves an oligonucleotide substrate in *trans*, allowing folding and cleavage steps to be studied separately and also allowing systematic variability of its catalytic components for easier dissection of the reactions steps. Since

only the native fold is able to cleave the substrate, ribozyme activity assays provide a sensitive readout for the fraction of ribozyme that has folded correctly [53].

The overall domain structure of the *Tetrahymena* ribozyme is separated into core and periphery [47] (Figure 1.4 B, C). The core helices P4-P6 and P3-P9 are highly conserved among other classes of group I introns and are stacked to form the active site cleft where guanosine binds and the P1 duplex docks. Peripheral elements vary in size and number among other group I classes. The periphery consists of P5abc, P2-P2.1 and P9-P9.1 which wrap around the core by using three long range tertiary contacts (P13, P14 and L9/P5) to form a peripheral ring. Two other tertiary contacts are formed along the P4-P6 domain with P5abc (A-rich bulge and tetraloop/tetraloop receptor). This interplay between the core and peripheral structure is critical for proper folding. Thus, the overall architecture proves to be tractable yet complex enough to have folding challenges associated with larger RNAs.

Taken together, this ribozyme is a good model system for RNA folding studies because: 1) folding and catalytic steps are distinct, 2) it provides a direct readout for the fraction of the RNA that has correctly folded, and 3) the relative size and complexity of this RNA is such that the associated folding challenge can be used as a model for other structured RNA's.

1.4.1 Folding of the *Tetrahymena* ribozyme

The folding of the *Tetrahymena* ribozyme has been extensively studied by Zarrinkar and colleagues using oligonucleotide hybridization to complementary sequences on the RNA [20, 32]. Results from these experiments suggested an initial

model for folding which included several intermediates. The first intermediate, which is populated in the absence of Mg^{2+}, consists of short-range secondary structure. Upon addition of magnesium, the RNA folds through another intermediate in which P4-P6 is formed, termed $I_{P4\text{-}P6}$ or I_{trap}. Finally, the rate limiting step was found to be arrangement of P3-P7 to another intermediate (I_3) before folding to the native state.

These studies were further supported by Sclavi *et al.* [9] who provided quantitative details on the folding kinetics of the ribozyme by monitoring the changes in solvent accessibility of each nucleotide at millisecond time intervals. In this study, they identified the fastest event as formation of P5a-P5c, which folded independently and constituted a Mg^{2+} rich location. This raised the possibility that the P5a-P5c region served as folding nucleation site, previously observed for some proteins. Next, tertiary interactions with P5abc and the P4-P6 domain were formed with a rate of about $1s^{-1}$. Together these results supported a model in which the formation of the P4-P6 and its tertiary interaction are critical for the folding of other domains.

The peripheral domains P2, P2.1, and P9.1 were protected from cleavage at rates of 0.2-0.4 s^{-1}, possibly reflecting the formation of tertiary contacts. The rate-limiting step, as previously suggested by Zarrinkar *et al.* [20], was folding of the P3-P8 domain and P9, indicating this region of the core remained disordered despite earlier formation of the long range native tertiary contacts. In general, studies on the folding of the ribozyme quantitatively detailed the overall hierarchy which occurs in the following order: 1) folding of the P5a-P5c sub-domain followed by rapid collapse of the P4-P6 domain, 2) inter-domain connection through long range peripheral contacts and 3) re-organization of the catalytic core by slow rearrangements in the P3-P7 domain

1.4.2 Misfolding in the *Tetrahymena* ribozyme

Early evidence of kinetically trapped intermediates in the *Tetrahymena* ribozyme came from oligonucleotide hybridization which revealed that the rate limiting step was folding of P3-P7. Furthermore, Woodson and coworkers used results from chemical modification interference assays and electrophoretic mobility assays to propose a model whereby the alternative base pairing of P3 to form Alt P3 was a feature of misfolding (Figure 1.5A) [54, 55]. Further insight into the misfolded intermediate came by utilizing a ribozyme activity assay which showed that 90% of the ribozyme misfolds at 25°C and that this misfolded state persisted on a time scale of hours [15]. SAXS experiments performed under the same experimental conditions as the activity assays revealed that the misfolded state was nearly as compact as the native state, with a slightly larger R_g value (51Å vs 47Å) [8].

A detailed study of the structural features of the misfolded RNA consisted of hydroxyl radical footprinting to compare nucleotide solvent accessibility between the native and misfolded states [35]. Differences in the P7, J3/4 and J6/7 regions confirmed that P3, not Alt P3 was formed in the misfolded structure. Since the region of misfolding was located in the core, it was reasonable that native contacts would have to come apart for the ribozyme to refold. As expected, disruption of tertiary contacts and adding urea accelerated refolding by 100-fold, indicating that non-native contacts are enforced by native contacts. Molecular modeling generated a feasible model that satisfied steric constraints by crossing strands at J3/4 and J6/7 (Figure 1.5 B). Therefore, a general model for misfolding in the *Tetrahymena* ribozyme predicted that the P3 helix was formed and also reconciled the earlier Alt P3 model by proposing that formation of Alt

P3 early in folding biases the ribozyme to misfold but is not present in the final misfolded structure (Figure 1.5C).

1.5 PERIPHERAL STRUCTURE IS CRITICAL FOR THERMODYNAMIC STABILITY OF THE NATIVE RIBOZYME

With the prevalence of highly structured intermediates and the propensity of RNAs to form misfolded structures, one challenge for structured RNA is to stabilize the native structure. For group I introns, peripheral elements have functional roles even though they are not generally conserved among intron subclasses. A well studied peripheral element, P5abc, from the *Tetrahymena* ribozyme was found to play a key role in stabilizing the native structure [56]. P5abc serves as a potential scaffold upon which further structure is built during folding, as this region forms early in folding along with the P4-P6 domain.

The functional roles of P5abc in folding and catalysis have been studied using a ribozyme variant that lacks P5abc ($E^{\Delta P5abc}$). Fe (II)-EDTA footprinting found that although this ribozyme is globally folded at 10-20 mM Mg^{2+}, catalytic activity, even at increased Mg^{2+} was severely compromised and could only become rescued upon the addition of P5abc. Supporting this result, addition of P5abc in *trans* formed a stable and tight complex [57, 58] and caused compaction of the core helices as observed in Fe(II)-EDTA footprinting. These results suggested a model in which P5abc plays a key role in catalytic activity by orienting the core helices within the active site [59]. Support for this model came from extensive kinetic and thermodynamic dissection of the catalytic steps

of $E^{\Delta P5abc}$ ribozyme, which showed that the absence of P5abc gave an astounding 10^4-fold decrease in overall catalytic activity of the ribozyme [60].

Peripheral structure in general can be one strategy for specifying the native structure by selectively stabilizing the native structure relative to alternative conformations. P5abc was demonstrated to display structural specificity by stabilizing the native state relative to a globally similar misfolded state, which is almost equally populated as the native ribozyme at equilibrium (K_{eq} = 1.4). Though the misfolded and the native state are globally similar with the same long range contacts, P5abc was found to selectively bind to the native ribozyme 50,000 fold tighter than the misfolded state. Together, these processes constituted a thermodynamic cycle (Figure 1.6), which indicates that P5abc indeed contributes to the thermodynamics of folding by shifting the equilibrium to favor the native state (K_{eq} = 70,000), thus conferring ~ 6 kcal/mol energetic stabilization [58].

Though the evidence presented above indicates that peripheral structure can provide thermodynamic assistance for stabilizing the native state, their role in folding kinetics has proved to be detrimental. This proverbial “catch 22” scenario is clearly demonstrated in studies with P5abc. This peripheral domain is the first to fold and forms a stable structure upon which the P4-P6 domain packs against. However, mutations that weakened P5abc interaction with the ribozyme increased the rate of formation for the stable structure [33, 61], suggesting that early formation of P5abc exacerbated folding to the native state [54, 62]. Further support for this unfavorable kinetic consequence arose in a quantitative study of folding of the $E^{\Delta P5abc}$ ribozyme. Time resolved footprinting of $E^{\Delta P5abc}$ revealed that the fastest folding events were located near the junction of P2 and

P2.1 and were at least four fold faster than the fastest folding events in the full length ribozyme. This result implied that early folding of P5abc in the full length ribozyme may inhibit structure formation within this domain. Interestingly, the overall folding to the native state and the partitioning ratio between the native and misfolded ribozyme $E^{\Delta P5abc}$ remained similar to the wild type. However, once the misfolded ribozyme was formed, the presence of P5abc hindered the refolding to the native state by nearly 100-fold.

We have seen here that P5abc, though detrimental to the folding kinetics, can provide a significant degree of thermodynamic stability to the native state. Figure 1.7 illustrates the stability of the unfolded, misfolded and misfolded states with the transitions between them. With P5abc, the energetic difference between than native and misfolded state is ~6 kcal/mol. P5abc can also slow the kinetics of folding to the native state, decreasing the energetic difference between the native and misfolded states. Overall, though peripheral elements can be detrimental in kinetics, they continue to be retained throughout evolution and in some cases, peripheral elements may be replaced by proteins which can function analogously to peripheral structure [63].

1.6 ALTERNATIVE MODE IN STABILIZING THE NATIVE STRUCTURE THROUGH THE ACTION OF RNA BINDING PROTEINS

In vivo group I introns are often associated with proteins suggesting that they may also be involved in stabilizing RNA structure. These proteins are separated into two groups based on whether they are encoded by the introns themselves or by the host [64]. Some group I introns contain an open reading frame that encodes for proteins that assist

in splicing. These proteins are termed maturases and may have originally functioned in intron mobility and possibly co-evolved with the RNA to function in splicing [64].

Host encoded proteins seem to have acquired splicing adaptations over evolutionary time. They include: leucyl-tRNA synthetase (NAM2) , CBP2 protein and the tyrosyl tRNA synthetase (CYT-18) [64]. The NAM2 protein assists in the splicing of bI4 and aI4α introns by acting in concert with a maturase [65, 66]. The yeast CBP2 protein assist in the splicing the bI5 intron. CBP2 has been shown to enhance the rate of splicing of bI5 intron by over 1000-fold [67]. The bI5 intron, however, folds differently from the *Tetrahymena* group I intron in that the bI5 intron collapses to a native-like intermediate upon the addition of Mg^{2+}. The lack of acceleration in folding by urea suggest that this native-like intermediate is not kinetically trapped [68, 69]. CBP2 binds stably to this native like intermediate in the region of P3, P8 and P7.1 and seems to capture native structure that forms transiently [70].

1.7 THE CYT-18 BINDING PROTEIN ACTIVATES SPLICING OF SEVERAL GROUP I INTRONS

The other group I intron splicing factor that is host encoded is the *Neurospora* mitochondrial tyrosyl tRNA synthetase (mt TyrRS) or CYT-18 protein, which has been the most studied and is the protein used in this study. CYT-18 catalyzes the aminoacylation of mt $tRNA^{Tyr}$ and also functions in the splicing of at least three introns in *Neurospora* mitochondria: 1) the large ribosomal subunit of rRNA intron (mt LSU), 2) the ND1 intron from the NADH reductase subunit I gene and 3) cob-I from cytochrome b gene [22, 71]. Additionally, purified CYT-18 has also been shown to interact with and

promote splicing of three non-cognate intron and ribozyme derivatives: the bI5 intron [70], $E^{\Delta P5abc}$ [63] and the *Twort* ribozyme [72]. The promiscuity of CYT-18 binding suggests that it recognizes structural features that are conserved among these RNAs.

CYT-18 protein functions as a homodimer, binding a single group I intron RNA [73]. Extensive structural probing with CYT-18 bound to either ND1 or mtLSU RNAs showed that CYT-18 binds along the P4-P6 domain on the opposite side from the active site cleft [21]. As a further test for these interactions, CYT-18 was able to compensate for mutational defects in J3/4, J6/7 and in P7 suggesting that it promotes the correct orientation of core helices [74, 75]. Though CYT-18 is homologous to the well studied bacterial synthetases, CYT-18, another mt synthetase from *P.anserina* and others closely related members in this protein family, have been found to function in group I intron splicing. Sequence alignments show that the CYT-18 protein has specific helical insertions in the nucleotide binding fold, Ins 1, Ins 2 and a N terminal extension H0 [76, 77] which have evolved to function in splicing (Figure 1.8 A).

Surprisingly, CYT-18 was demonstrated to function similarly to the peripheral structure P5abc by promoting splicing, even at low Mg^{2+} concentration. Indeed, structural probing provided support for a critical role in stabilization of the native structure since CYT-18 dependent protections on the RNA were observed along J3/4, J4/5, J5/4, J6/6a, J7/3, J8/7 and L9, even at low Mg^{2+}. These protections were not as strong as in the presence of P5abc or for the full length ribozyme, suggesting that CYT-18 did not fully compensate for P5abc as would be reasonable since the P5abc peripheral element co-evolved with the RNA [63]. This discovery forms the foundation on this work herein and

provided insight into how an RNA binding protein could replace peripheral structure in folding and activity.

1.8 CYT-18 SPECIFIC INTERACTIONS WITH GROUP I INTRONS.

The high resolution crystal structure of the C-terminal deleted CYT-18 (CYT-18Δ424-669) at 1.95Å highlights the three helical insertions (Figure 1.8B). Using experimental constraints from directed hydroxyl radical cleavage, CYT-18 was docked onto both $tRNA^{Tyr}$ and ND1 intron, revealing that the intron did not occupy the tRNA binding site as previously suggested, thus providing evidence that CYT-18 evolved a new binding interface for group I intron splicing [77].

The 4.5Å co-crystal structure of CYT-18/Δ 424-669 bound to the *Twort* ribozyme provided additional details into CYT-18 specific interactions important for stabilizing the native structure [78]. CYT-18 insertions are positioned to interact with the P4-P6 domain to fully buttress the domain to promote its correct orientation (Figure 1.9). The structure also highlights how CYT-18 stabilizes the two core domains. CYT-18 contacts the single-stranded region in the core that connects the P4-P6 domain to the P3-P9. Helix H0 interacts with J3/4 to force J3/4 into the P6 minor groove for a triple helical interaction. Another core strand J6/7, forms a triple helical interaction with P4 and also interacts with P7 to form the guanosine binding site, thus helix H0 plays a critical role in the RNA core interactions. CYT-18 uses Ins2 to “cap” the interaction between the L9 tetraloop and the minor groove of P5. This interaction may serve as the top anchor point for correct orientation of P4-P6 with the P3-P9 domain.

The majority of CYT-18's interactions are with the RNA phosphodiester backbone, which explains why CYT-18 is able to bind to various group I introns. A visual comparison of CYT-18 binding compared to its RNA counterpart P5abc provides answers to many questions concerning the ability of CYT-18 to functionally replace P5abc. Though the physical nature of RNA/RNA contacts and RNA/protein contacts differ, CYT-18 has evolved to make contacts along the P4-P6 domain, acting as a molecular scaffold for the correct orientation of the P3-P9 domain to the P4-P6 domain.

1.9 OVERVIEW OF DISSERTATION RESEARCH

CYT-18 protein and the $E^{\Delta P5abc}$ ribozyme, a model system for mechanistic studies in protein assisted RNA folding

The work contained in this dissertation is aimed at understanding the mechanism of protein assisted RNA folding. Using CYT-18 and the $E^{\Delta P5abc}$ ribozyme, I set out to probe the kinetic and thermodynamic contributions of preferential stability of the native ribozyme relative to a misfolded ribozyme. Previous work in our lab indentified a misfolded $E^{\Delta P5abc}$ ribozyme that persists at equilibrium, in almost equal amount as native ribozyme [58]. In order to fully understand CYT-18 interactions with the misfolded ribozyme, it was first necessary to characterize its structure. In chapter 2, I extensively probed the structural features of this misfolded $E^{\Delta P5abc}$ ribozyme and found that it was highly similar to the misfolded full length ribozyme that was previously characterized [35]. Chapter 3, represents the core of my dissertation work. I quantatativly measured the extent of the thermodynamic stability afforded by CYT-18 to the native state relative to the misfolded state and also probed its overall effects on the kinetics of folding to the

native state. I was able to report an important feature of protein assisted RNA folding, in that CYT-18 conferred thermodynamic stability to the native state without compromising the ribozyme folding kinetics. This was a stark contrast to P5abc where it dramatically slowed the kinetics of folding [79]. In chapter 4, I probed the mechanistic origin of CYT-18 stabilization of the native ribozyme by disrupting one long range contact P14, which forms part of a complete ring of contacts around the core. Though both CYT-18 and P5abc occupy the same binding interface on the RNA, formation of the P14 contact was found to be critical for P5abc; however not for CYT-18.

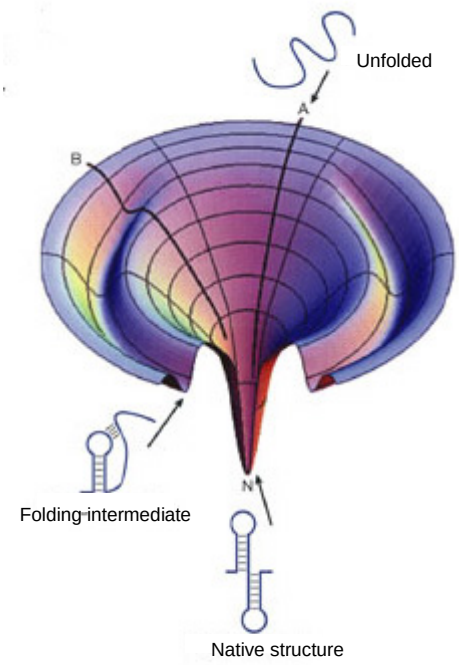

Figure 1.1: Representation of the folding landscape traversed by a structured RNA

Radial lines around the folding landscape represent any possible RNA conformer that can exist upon initiating folding. The vertical axis which lead into the deep tunnel represent the internal free energy of each conformer. Pathway A is folding of an RNA molecule directly to the native state with no significant long-lived kinetic traps. Pathway B goes through at least one kinetically trapped intermediate. The deep valley represents energetic barriers in which kinetically trapped intermediate must be over come to fold to the native state [23].

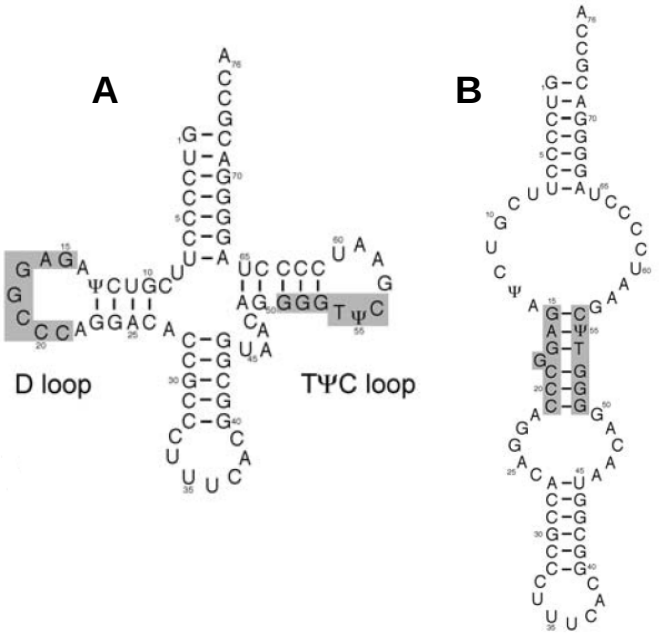

Figure 1.2: Native and misfolded forms of tRNA

A) Canonical clover leaf shape of native E.coli $tRNA^{Glu}$ **B)** Misfolded RNA is shaped as a hairpin with two single stranded bulges. Non -native base pairs are formed between nucleotides in the D loop and TΨ C loops [25] [34].

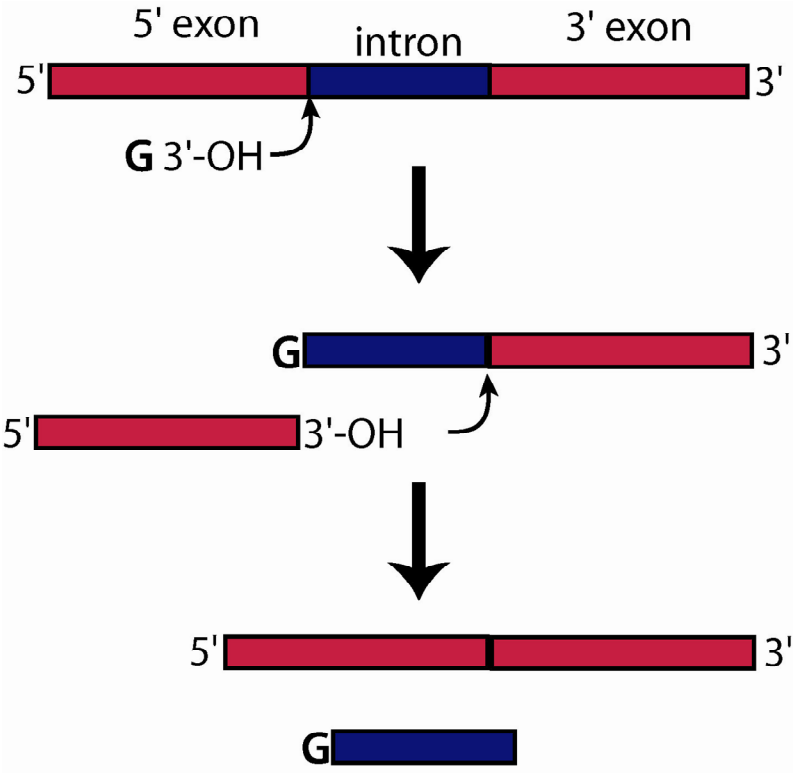

Figure 1.3: Schematic of group I intron splicing

Exon sequences are shown in red and the intron is shown in blue. The 3’ OH of an exogenous guanosine nucleotide attacks the 5’ exon resulting in strand cleavage. The 3’OH of the excised 5’exon next attacks the 3’splice site. The final splicing products are the excised intron and the ligated exons.

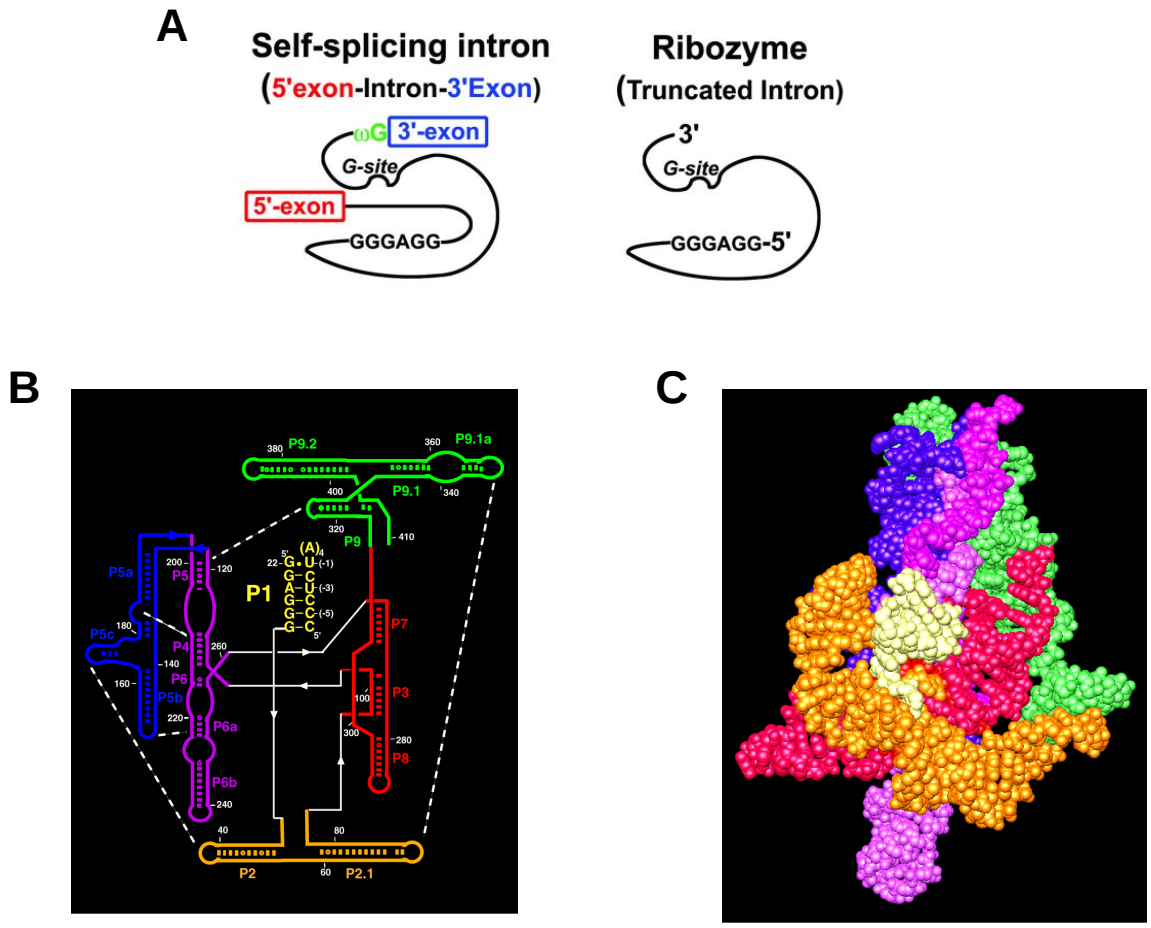

Figure 1.4: The global structure of *Tetrahymena* ribozyme

A) Conversion of the *Tetrahymena* group I intron to a ribozyme by removal of 5' and 3' exons [52]. **B)** Secondary structure map of the *Tetrahymena* ribozyme showing paired helices (P). Core domains are shown in purple and red and peripheral domains in blue, yellow and green. Five tertiary contacts are shown in white. The P1 duplex is formed by base pairing with substrate. Cleavage of the substrate produces products 5'CCCUCU 3' and 5' GAAAAA 3'. **C)** Space filling model of the globally folded ribozyme color coded as in B. Peripheral elements in yellow and green wrap around the core by forming a circle of contacts.

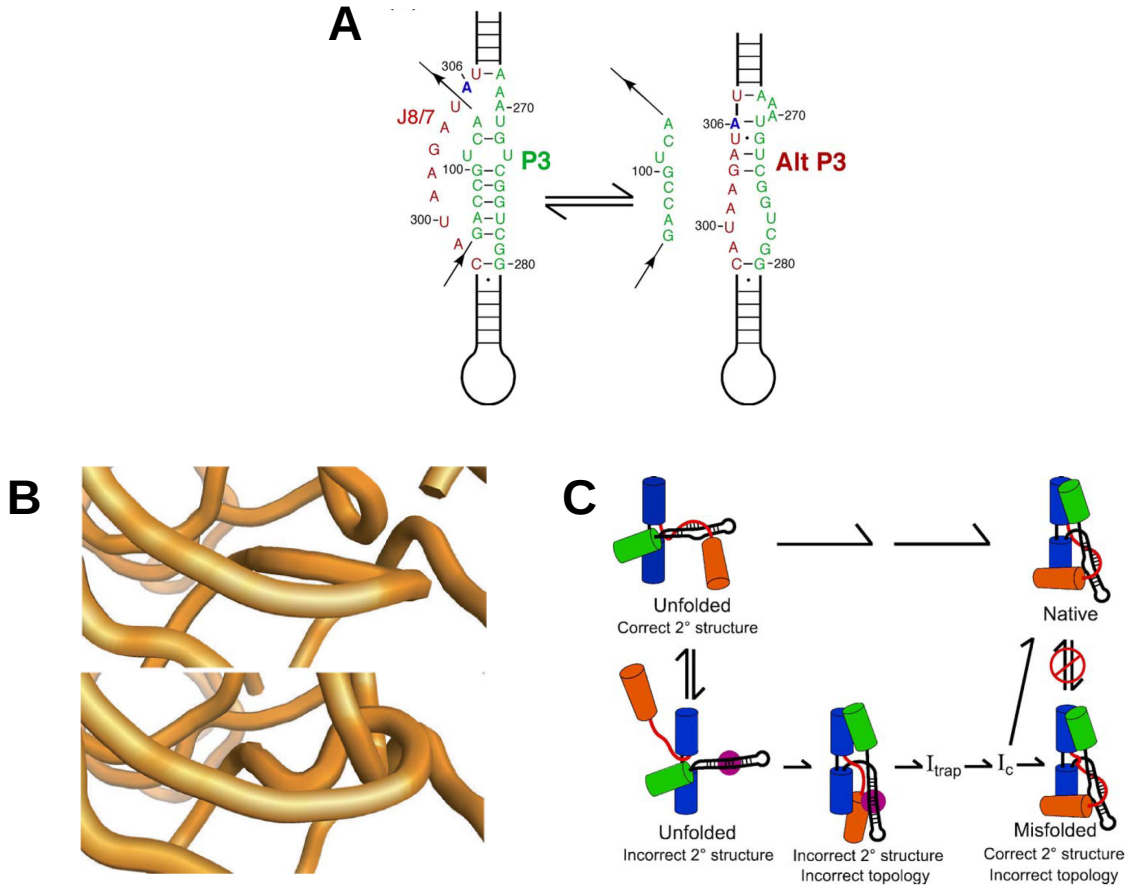

Figure 1.5: Alternative base pairing in the *Tetrahymena* ribozyme

A) Native P3 base pairing is shown in green. When nucleotides in J8/7 participate in base pairing this helix is referred to as AltP3. **B)** Molecular modeling of strand crossing at C260 and U305. **C)** Model for the formation of the long lived misfolded intermediate, a topological isomer. The cylinders are P4-P6 (blue), P2 (orange) and P9 (green). The P3-P8 helix and the 5'portion of P3 are shown as black and red strands respectively. Other helices are omitted for clarity. The correct secondary structure gives a folding pathway directly to the native state (top). The presence of Alt P3 (purple ball) biases folding to the long lived misfolded state. The misfolded structure must unfold and re-fold back to the native state [35].

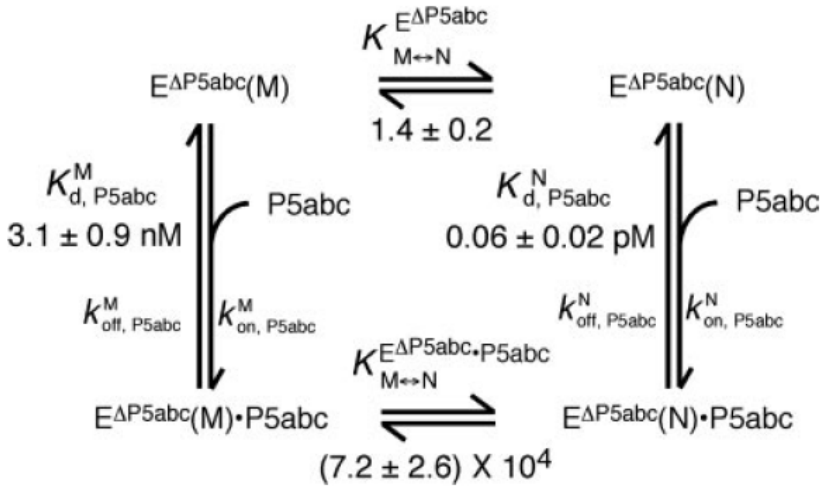

Figure 1.6: Relative stabilization of P5abc to the native ribozyme

The relative stabilization afforded by P5abc was calculated by the differences of P5abc binding to the native (N) and misfolded (M) ribozymes. P5abc bind 50,000 fold tighter to the native state relative to the misfolded state. Since deletion of P5abc gave a equilibrium constant near unity (K_{eq}=1.4), the relative stabilization afford by P5abc was calculated to be ~ 72,000 [58].

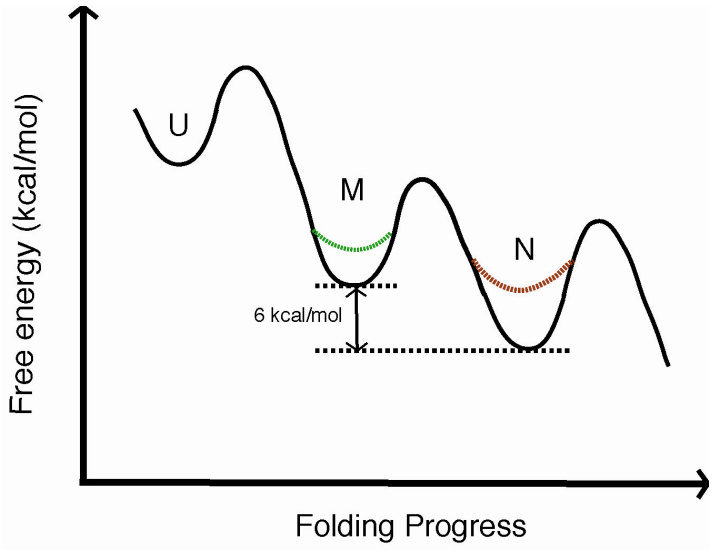

Figure 1.7: Free energy profile of ribozyme folding

Diagram shows the unfolded (U), misfolded (M) and native (N) ribozymes and their folding transition states. Red dotted line represents the destabilization of the native ribozyme upon removal of P5abc. Black dotted lines show the relative stability (6 kcal/mol) conferred to the native state by the P5abc peripheral element. The green dotted line represents the decrease in the energetic difference in M thus representing a smaller energetic “hill” during refolding to the native state.

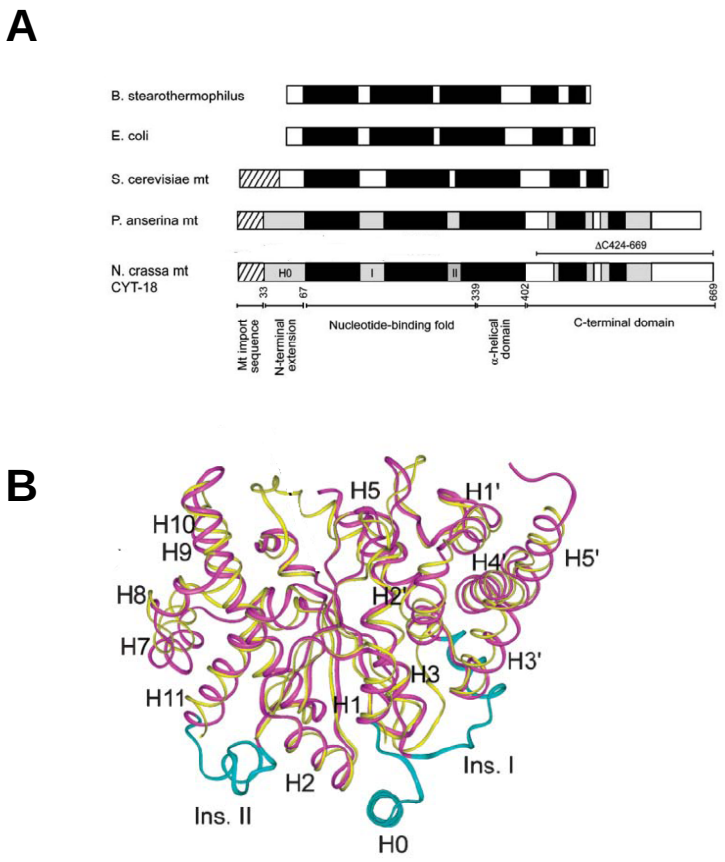

Figure 1.8: Domain comparison of CYT-18 with bacterial and mt-tyrosyl tRNA synthetases and identification of CYT-18 helical insertions

A) Bar representation of nucleotide binding fold, α helical domain and C-terminal domain of bacterial and mitochondrial synthetases. Regions of total conservations are shown in black. The mt import sequence is shown with diagonal lines. Helical insertions present only in CYT-18 and *P.anserina* are shown in grey. They are H0, I and II. **B)** Ribbon diagram of CYT-18/Δ424-669 in purple and *B.stearothermophilus* (yellow) TyrRS superimposed on each other. These highlight the helical insertion (blue) that are only present in CYT-18 and *P.anserina*. These insertion have evolved to allow the protein to function in group I intron splicing [77].

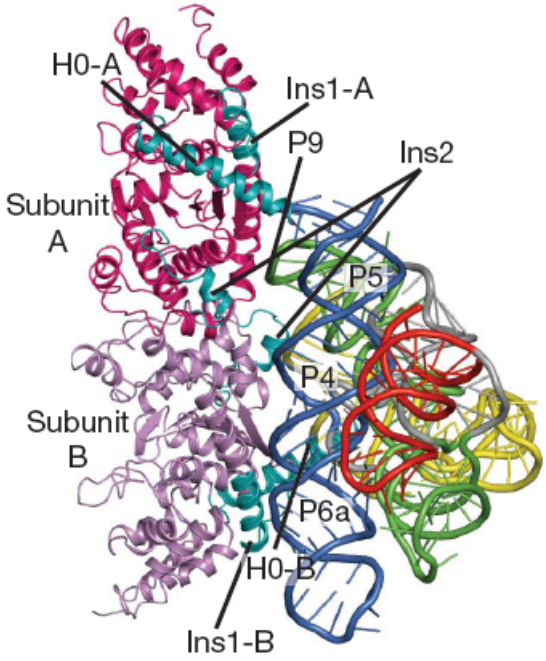

Figure 1.9: Co-crystal structure of the CYT-18 and *Twort* ribozyme

The co-crystal structure was determined at 4Å resolution. CYT-18's homodimer subunits are colored magenta and violet. CYT-18's helical insertions, H0, InsI and Ins2 on each subunit are shown in cyan. CYT-18 helical insertions make specific contacts also the P4-P6 domain (blue) of the *Twort* ribozyme. Domains in the *Twort* ribozyme are colored as follows P2 (red), P3-P8 (green) and P7.2/7.2/9.1(yellow) [72].

Chapter 2: Structural probing of the native and misfolded conformations of the P5abc deletion variant from the Tetrahymena ribozyme

2.1 INTRODUCTION

Structured RNAs must fold into specific three-dimensional structures, the complexity of which is dictated by their functional roles in the cell. The multiple steps an RNA takes to go from a linear molecule to a globally compact structure are reflected in a series of folding intermediates, each accumulating more structure than the intermediate before, likened to a general hierarchal assembly [8, 23]. These intermediates do posses some level of structure as RNA helices are independently stable even in the absence of tertiary contacts. Given that these folding intermediates have a high degree of stability, how then are RNAs able to specify its native structure relative to all other structures populated along the folding pathway?

Initial studies using the *Tetrahymena* ribozyme explored the contribution of peripheral helices to the stability of the native state. The effect on native stability has been most extensively studied using the variant ribozyme lacking P5abc ($E^{\Delta P5abc}$) by monitoring the biochemical and structural changes upon adding back P5abc as a separate molecule. It was found that P5abc contributed to catalysis and was necessary for promoting the formation of the native state [15, 56, 80, 81]. Furthermore, the effect of P5abc on stabilizing the native state was found to be specific in that P5abc bound 50,000 times tighter to the native state relative to a defined misfolded state, giving 6 kcal/mol of stabilization [58].

Like peripheral elements, proteins also can play similar roles in stabilizing the native state and promoting catalysis [64]. *In vivo* most group I RNAs are closely associated with proteins, suggesting that they are directly involved in folding and promoting catalysis of RNAs. For example, the CYT-18 protein associates with several group I introns by binding along the P4-P6 domain, a region occupied by the P5abc peripheral element in the *Tetrahymena* ribozyme [82, 83]. In further support for this, $E^{\Delta P5abc}$ ribozyme and CYT-18 were previously shown to form a functional and catalytically active complex, demonstrating that CYT-18 could indeed function similarly to P5abc [63].

While it is clear that CYT-18 binding provides assistance in stabilizing the native state, it is not known to what extent it has evolved to confer stability to the native state relative to the next most stable RNA folding intermediate. The result observed that $E^{\Delta P5abc}$ ribozyme populated a long-lived misfolded state at equilibrium [58], provided us with an experimental advantage in probing the extent through which a protein can stabilize the native state relative to the next most stable state.Therefore, to investigate if CYT-18 is able to preferentially stabilize the native state relative to a defined misfolded state, it was first necessary to ensure that the ribozyme was globally folded under our experimental conditions and furthermore, to characterize the misfolded ribozyme to see to what extent, if any, it differed from the misfolded state of the full length ribozyme.

Hydroxyl radicals are useful chemical probes to measure solvent accessibility of each nucleotide without sequence or secondary structure specificity [84]. Hydroxyl radical footprinting was successfully used to distinguish the differences between the native and misfolded folds in the full length *Tetrahymena* ribozyme [35]. We therefore

used this method to probe the native and misfolded states for $E^{\Delta P5abc}$ ribozyme. We first followed the equilibrium folding of $E^{\Delta P5abc}$ to ensure that the ribozyme was globally folded under our standard experimental conditions. Next, we generated population of native and misfolded ribozyme in the presence and absence of P5abc to probe changes in nucleotide solvent accessibility. These results revealed that regions of differences between the native and misfolded $E^{\Delta P5abc}$ are very similar to the full length ribozyme, but with some local differences attributed to rearrangements upon P5abc binding. [58, 79]

2.2 MATERIALS AND METHODS

2.2.1 Preparation of RNA

DNA encoding the P5abc deletion variant ribozyme ($E^{\Delta P5abc}$) and the P5abc element was cloned into a pUC18 plasmid. In vitro transcription was carried out either from a Sca I (New England Biolabs) for $E^{\Delta P5abc}$ or Bsa I (New England Biolabs) for P5abc digested plasmids [15, 51]. RNAs were transcribed under the following conditions: 5 μg/ml linearized vector, 7.5 μg/ul, T7 RNA polymerase (0.3 mg/ml), 1 mM ATP, UTP, GTP and CTP, 40 mM DTT and 1X transcription buffer (40 mM Tris.HCl, pH 8.0, 25 mM $MgCl_2$, 2 mM spermidine and 0.01% Triton X-100). *In vitro* transcription was carried out at 37 °C for 4 hours after which RNA was precipitated overnight at -20 °C in 1/10 V 0.5 M EDTA pH 8, 3V 100% ethyl alcohol, 1/10 V 3M sodium acetate and 2M acetic acid. RNAs were purified using the manufacture's protocol from the Qiagen RNeasy kit and stored in 10 mM Tris, 1 mM EDTA at -20 °C.

2.2.2 End-labeling of RNAs for footprinting studies

RNAs were 5' end-labeled by the removal of 5' phosphoryl groups using shrimp alkaline phosphatase (Fermentas) at 37 °C for one hour. After inactivation of the phosphatase (65 °C for 15 minutes), [γ-^{32}P] ATP (Perkin Elmer, 5 mCi stock,) and T4 polynucleotide kinase (Fermentas) were added for a one hour incubation at 37 °C [51]. Reactions were stopped using 20 mM EDTA, 0.01% bromophenol blue and 0.01% xylene cyanol in 95% formamide and purified on an 8% native polyacrylamide gel.

RNAs were 3'end-labeled using the procedure essentially as described from Huang and Stosack [85]. DNA oligonucleotide is annealed to the 3' end of the ribozyme with a two nucleotide overhang 5'-GTA CTC CAA AAC TAA TCA ATA TA/3ddC-3'. The DNA oligonucleotide is designed to have a 2' 3' dideoxyC to prevent its unnecessary labeling. 10 μM of a DNA oligonucleotide is annealed to the ribozyme in the presence of 1X annealing buffer (14 mM Tris-HCl pH 7.5, 40 mM NaCl, 0.2 mM EDTA) at 65 °C for 1 minute. To extend the RNA, a Klenow fragment (3' → 5' exo) (500 U/ml (final), New England Biolabs) and [α-^{32}P] dATP (Perkin Elmer, 20 μCi/μL) was added to the reaction on the presence of a 1X Mg^{2+} buffer solution (7 mM $MgCl_2$, 1 mM DTT). The reaction was allowed to proceed for 2 hours at 37 °C. The total volume (100 μL) was concentrated to 15-20 μl using a Micron YM-50 (Millipore) and next loaded onto a P-30 Bio-Spin column (Bio-Rad) to remove unincorporated nucleotides. Reactions were stopped using 20 mM EDTA, 0.01% bromophenol blue and 0.01% xylene cyanol in 95% formamide. Ribozymes were then purified on an 8% polyacrylamide, 7M urea gel. Bands were excised and RNA eluted in 10 mM Tris-HCl, pH 8.0 and 1 mM EDTA overnight at 4 °C and stored at -20 °C at ~ 100 000 cpm/μl.

2.2.3 Equilibrium footprinting of $E^{\Delta P5abc}$ using Fe(II)-EDTA

5’ or 3’ end-labeled $E^{\Delta P5abc}$ (20,000 cpm/µl) was folded in Na-MOPS, pH 7.0 buffer at various Mg^{2+} concentrations (0–50 mM) for 2 hours at 25 °C to allow equilibration. Next, 1/10 reaction volume of Fe(II)-EDTA footprinting reagent was added (2 mM $(NH_4)_2Fe(II)(SO_4)_2$, 2.5 mM Na-EDTA, 6 mM Na-ascorbate) and incubated for 10 minutes at 25 °C. Reactions were stopped in one-half reaction volume of Fe(II)-EDTA quench solution (33 mM thiourea, 0.03% bromophenol blue, 0.03% xylene cyanol and 30 % formamide) [58]. All samples were resolved on either an 8% or 18% denaturing PAGE that was initially ran for 0.5-2 hours at 55 Watts in 1X TTE buffer (70 mM Tris, 20 mM taurine, 0.4 mM Na-EDTA). An RNA sequencing ladder was generated by RNA digestion using RNase T1 under denaturing conditions (7M urea) to identify all the guanosine nucleotides.

2.2.4 Footprinting of the native and misfolded states of $E^{\Delta P5abc}$ either alone or in complex with P5abc

5’ or 3’end-labeled ribozymes were folded to the native state by incubating in 50 mM Na-MOPS buffer, 10 mM Mg^{2+} at 50 °C for 30 min [15]. 200 nM P5abc was included in this incubation for footprinting the native complex with P5abc [15, 57]. The misfolded ribozyme was generated under the same solution conditions, except that it was incubated at 25 °C for 10 min. P5abc was added for an additional 5 minutes at 25 °C to trap the misfolded ribozyme. The misfolded intermediate generated here was shown to be the same misfolded intermediate which persists at equilibrium [15, 62]. For all reactions, footprinting reagent was added and reactions were quenched as described above.

2.2.5 Quantitative analysis of footprinting bands using Semi-Automated Footprinting Analysis (SAFA) software.

This method has been published by Das et al. in RNA [86]. SAFA software is readily available at (http://safa.stanford.edu) free of charge and is executable on MATLAB and windows platforms. SAFA allows for rapid gel analysis achieved through a series of commands which allows: lane definition, gel alignment and assignment of residue number. A great advantage of this software, besides its speed of quantitation, is that it is designed for users to refine guesses made by the program in defining the lanes of the gels or in assigning nucleotide numbers to the bands. Each gel is uploaded and footprinting bands are assigned by its sequence number. Bands are then fit to the sum of Lorentzian peak areas which gives the number of counts in each band. Data for each nucleotide is then exported to an Excel file where band intensities are normalized to the total averaged band counts in each lane within the region of analysis.

2.3 RESULTS

2.3.1 $E^{\Delta P5abc}$ equilibrates almost equally between the native and misfolded state

Although the full length ribozyme initially misfolds, the native state is favored at equilibrium such that refolding from the misfolded ribozyme is essentially complete. Interestingly, removal of the P5abc peripheral domain destabilized the native state such that at equilibrium, the ribozyme populated the native state with a much lower endpoint than observed for the full length ribozyme (0.47 vs 0.9). The endpoint for the full length never reaches 1 due to some fraction of damaged ribozyme (~10-20%) within each RNA preparation. The lower endpoint for $E^{\Delta P5abc}$ could be interpreted as the presence of a) presence of a misfolded state b) more damaged ribozyme, c) a kinetically trapped intermediate or d) a new equilibrium between the native and partially folded intermediates.

Experiments were performed to test which of these possibilities held true. The presence of damaged ribozyme or the presence of a kinetically trapped intermediate was ruled out by T. Johnson using ribozyme activity assays. Adding back P5abc in *trans* to $E^{\Delta P5abc}$ produced a much higher endpoint than the equilibrium endpoint confirming that the rest of the ribozyme was not damaged. Next, to rule out possibility of a kinetically trapped intermediate, equilibrium folding was monitored under two different Mg^{2+} conditions, first at 10 mM Mg^{2+}, then a jump to 50 mM Mg^{2+}. The ribozyme approached a new and higher endpoint when monitored at 50 mM Mg^{2+}, indicating a higher fraction of native ribozyme. As expected, dilution back to 10 mM Mg^{2+} led to the return to the initial fraction of native ribozyme present at equilibrium (0.47) [58]. This result is not

indicative of a kinetically trapped species because a kinetically trapped species would be less stable than the native species and would not be able to increase in concentration relative to the native species.

To address the last possibility in which there was a new equilibrium between the native and partially folded intermediates, I probed the stability of the RNA under varying magnesium concentrations using Fe(II)-EDTA footprinting under the same experimental conditions used in activity assays to monitor equilibrium folding (25 °C, 50 mM Na-MOPS, pH 7). Hydroxyl radicals generated by the Fenton reaction were carried out at 25°C and stopped using a thiourea quench [84]. Figure 2.1 shows a representative schematic of probing RNA structure by hydroxyl radical footprinting. Nucleotides that are exposed will have a higher probability of undergoing strand cleavage than those nucleotides that are protected by other RNA helices.

$E^{\Delta P5abc}$ was equilibrated for 2 hours at 25 °C at each Mg^{2+} concentration, a sufficient time needed to achieve equilibrium under the range of magnesium concentrations used [58, 87]. Figure 2.2 shows a representative gel of cleavage bands from a 3' end-labeled RNA. We observed clear protections and enhancements in regions highlighted with blue and red bars respectively. "Protections" suggest that the RNA backbone is becoming increasingly solvent inaccessible due to compaction of helical elements or possibly from the formation of tertiary contacts. "Enhancements" refers to the region of the backbone that may have started out as being protected but now becomes exposed to the solvent as the ribozyme folds. Each band was quantified using Semi-Automated Footprinting Analysis software (SAFA) [86] and normalized to the average intensity of all bands in each lane within the region of SAFA analysis. Protections and

enhancements induced by Mg^{2+} were considered to be significant if they were observed in each of four independent experiments and the band intensity value changed by a value at least 30% of the background minus the signal obtained in the unfolded ribozyme.

Overall the data showed that the ribozyme achieves its global fold with a $K_{1/2}$ value of ≤ 5 mM Mg^{2+}. Many of the protections observed were in the peripheral helices P2.1, P2, P8, P9.1 and P9.1a and are similar to those previously observed in a study by Doherty *et al* (Figures 2.3 and 2.4). Interestingly, we observed several regions of enhancements which had not been reported for this ribozyme (Figure 2.3, red bars). Taken together, $E^{\Delta P5abc}$ ribozyme is fully folded under the conditions of activity assays and the mixed population seen at equilibrium using the ribozyme activity assays can now definitively be identified as between the native state and some long-lived misfolded intermediate and not between the native state and a partially folded or unfolded species. Furthermore, footprinting gave a good indication of what parts of the ribozyme gave full protection with the lowest magnesium requirements (Table 2.1).

2.3.2 Structural mapping of the native and misfolded ribozyme of $E^{\Delta P5abc}$

P5abc clearly contributes to the thermodynamic stability of the native state, as deletion of this domain led to almost equal population of the native state and a long-lived misfolded state at equilibrium. Here, we wanted to probe the structural characteristics of this misfolded state to see if the nature of the misfolding differed from the full length ribozyme. The structural features of the misfolded ribozyme of the full length were probed by hydroxyl radical footprinting and DMS footprinting. In addition, its urea dependence of refolding and accelerated refolding upon tertiary contact mutations all

provided evidence that the misfolded ribozyme is globally similar to the native ribozyme, with the same long range tertiary contacts. The main difference between the native and misfolded ribozyme was localized to a region of the core and suggested a topological model whereby strands J3/4 and J6/7 are crossed incorrectly [35].

As with the full length ribozyme, structural probing using Fe(II)-EDTA footprinting was used to distinguish the native and misfolded state for the $E^{\Delta P5abc}$ [55, 79]. Populations of either the native or misfolded ribozymes were formed, after which hydroxyl radical cleavages were initiated by the addition of footprinting reagent at 25 °C. Cleavage bands were visualized by using a phosphorimager and a representative gel is shown in Figure 2.5. Lanes 1-6 show the footprinting bands for a region of ribozyme spanning nucleotides 240-360. Footprinting patterns for the native ribozyme (lane 1) were first compared to the ribozyme in the absence of magnesium (lane 3), to ensure that the ribozyme obtained a global fold under our footprinting conditions. There were several regions of protections in P7, P3, P4, P2 and P9 and enhancements observed in P6b, L2.1, L9 and L8 (See Figure 2.7). While there were protections between unfolded and fully folded ribozyme (U vs N), we observed no significant differences in solvent protections between the native and misfolded ribozyme as observed on the gel. Quantitation using SAFA confirmed that there were no significant differences (Figure 2.6, thin red, blue lines).

We reasoned that the relatively weak protections and enhancements were a reflection of a more dynamic structure for $E^{\Delta P5abc}$ as also observed previously by Doherty *et. al*. [57]. Another explaination is that the native population generated for $E^{\Delta P5bac}$ was a heterogeneous mixture, containing significant population (~ 40%) of misfolded ribozyme

at equilibrium [58]. This heterogeneity would lessen any signal obtained for the native state. The results presented here highlight some limitations that prevented initial analysis of differences between the native and misfolded ribozymes.

2.3.3 Structural mapping of $E^{\Delta P5abc}$ with P5abc added in *trans*: The ribozyme deletion variant ribozyme populates a similar misfolded intermediate as the full length ribozyme

The observation that there were at most small differences between the native and misfolded $E^{\Delta P5abc}$ ribozyme suggested two experimental limitations listed above. In this section, we address these limitations through the addition of P5abc in *trans*. Adding P5abc to a population of misfolded ribozyme essentially "traps" the misfolded RNA thus preventing re-folding from occurring [15, 79]. P5abc was also added during the incubation time necessary to generate the native ribozyme, serving to increase the population of native ribozyme formed in order to obtain a greater footprinting signal. Lanes 2 and 6 of Figure 2.5 show the footprinting pattern of the native and misfolded ribozymes with P5abc added as a separate molecule. Both complexes gave larger signals in footprinting, so it was now possible to observe differences between the native and the misfolded RNA (Figure 2.5 black bars).

Each band on the gel was quantified using SAFA and band intensities were normalized to its average lane intensity. Plots showing these normalized values are shown in Figure 2.6. We observed clear differences between the native (thick blue line) and misfolded ribozymes (thick red line) which were also seen in the full length, for example in residues 264-270 and 307-310 (Figure 2.6, black boxes) [35]. We also

observed some small differences between the native and misfolded ribozymes which were not detected for the full length ribozyme. Nucleotides in P5 and J8/7 show more protection in the native state. These protections have probably arisen by structural rearrangements when P5abc was added back. P5 is the pivot point of a covalent contact between P5abc to the P4-P6 domain in the full length ribozymes suggesting that protections observed in the native can be explained by formation of the *trans* complex which are not possible when P5abc is not covalently attached. The lack of these protections without P5abc would support this. A second difference was seen in J8/7. It is possible that these additional differences may arise from small rearrangements that occur upon P5abc binding misfolded ribozyme. A slightly more dynamic nature of the misfolded RNA in complex with P5abc complex would give localized regions of increased solvent accessibility that are not present in the native state. The 50,000 fold difference in affinity for P5abc to the native state relative to the globally similar misfolded state would support the changes observed which can explain the differences in the binding interface [58]. Thus, these results from extensive structural probing have highlighted the features of the misfolded intermediate of $E^{\Delta P5abc}$ which are highly similar to the misfolded conformation of the full length ribozyme.

2.4 DISCUSSION

RNAs are diverse and ubiquitous molecules that function in an enormous variety of biological process such as gene regulation [36, 88, 89] and translation [90, 91]. To function, all structured RNAs must adopt a specific three dimensional structure. The folding process traversed by each RNA molecule is aimed at attaining a specific three

dimensional fold that must overcome stable and long-lived intermediates to specify a single functional structure [7]. Group I introns have been an invaluable model in studying RNA stability. One strategy these RNAs have adopted to stabilize the active structure is through the use of RNA peripheral elements [92].

One of the more extensively studied peripheral elements is P5abc from the *Tetrahymena* ribozyme. Deletion of P5abc gave a considerable decrease in activity [60] and in the fraction of native ribozyme populated at equilibrium (endpoint of 0.47) [58]. With such a low endpoint, the question arose as to what was the identity of the remaining population of ribozyme present at equilibrium. This population likely reflects the presence of a misfolded intermediate, however other possibilities could potentially hold true. These possibilities include the presence of dead or damaged ribozyme, the formation of a kinetically trapped intermediate or some new equilibrium between the native state and some partially folded intermediate.

Here we specifically address the possibility in which a new equilibrium was established between the native state and a partially folded state. Though previous experiments [57] indicated that the ribozyme attained a global fold between 5-10 mM Mg^{2+}, the experimental conditions were different to what we used in the ribozyme activity assays. It was thus necessary to ensure that the ribozyme attained a global fold under conditions in which the approach to equilibrium was measured in the ribozyme activity assays. The folding intermediates were monitored under equilibrium conditions with concentrations of Mg^{2+} ranging from 0-50 mM. Probing the solvent accessibility at a nucleotide resolution has an advantage because there is a general idea of the global structure of any possible intermediates. Fe(II)-EDTA footprinting gave protections in

P2.1 and P9 and several regions of enhancements were located on P6b and P9, P2 and P2.1. Most of the observed changes in accessibility were seen with $K_{1/2}$ of 1 mM Mg^{2+}, providing evidence against a partially folded intermediate.

In addition, the stability of $E^{\Delta P5abc}$ at varying concentrations of Mg^{2+} by ribozyme activity assay gave a $K_{1/2}$ of 2.4 mM Mg^{2+}, confirming that the RNA attains its global structure [79]. The full length ribozyme is very stable, with a $K_{1/2}$ value of 0.4 mM Mg^{2+}, supporting the result that P5abc is critical for stability of the native fold. Interestingly, the stability of the full length ribozyme with disruption of a tertiary contact made with P5abc (P14) gave $K_{1/2}$ value of 1.8 mM (H.Bhaskaran and I Jarmoskaite), and a ribozyme with a disruption of another contact made by P5abc, L5b, gave a $K_{1/2}$ =1.2 mM, both relatively close to the $K_{1/2}$ value for $E^{\Delta P5abc}$, suggesting that tertiary contacts made by P5abc are critical and independently contribute to the stability of the native ribozyme.

The result that $E^{\Delta P5abc}$ is globally folded at 10 mM Mg^{2+} and populates both the native state and a misfolded state at equilibrium led to the next logical step; the characterization of this misfolded ribozyme. The native and misfolded ribozyme were probed by hydroxyl radical footprinting which surprisingly showed no large differences in footprinting pattern. This was probably a result of a less compact global structure [57], or presence of a heterogeneous population of the native and misfolded ribozymes. These limitations were overcome by the addition of P5abc in *trans* such that differences in footprinting pattern between the native and misfolded were now observed. We noticed the same differences as previously observed in full length ribozyme suggesting the nature of the misfolding to be the same. Additional differences were also observed in P5, J8/7

and P3 and are probably due to local changes in solvent accessibility from P5abc binding in *trans*.

Taken together, extensive footprinting analysis of the $E^{\Delta P5abc}$ ribozyme was performed which confirmed that the ribozyme attains a global fold with an overall $K_{1/2} \leq$ 5 mM for magnesium dependent folding, with several of the peripheral helices becoming protected with $K_{1/2} \leq 1$ mM Mg^{2+}. In addition, we have fully characterized the structural differences between the native and misfolded ribozyme and concluded that the nature of the misfolding for $E^{\Delta P5abc}$ is fundamentally the same as that of the full-length ribozyme.

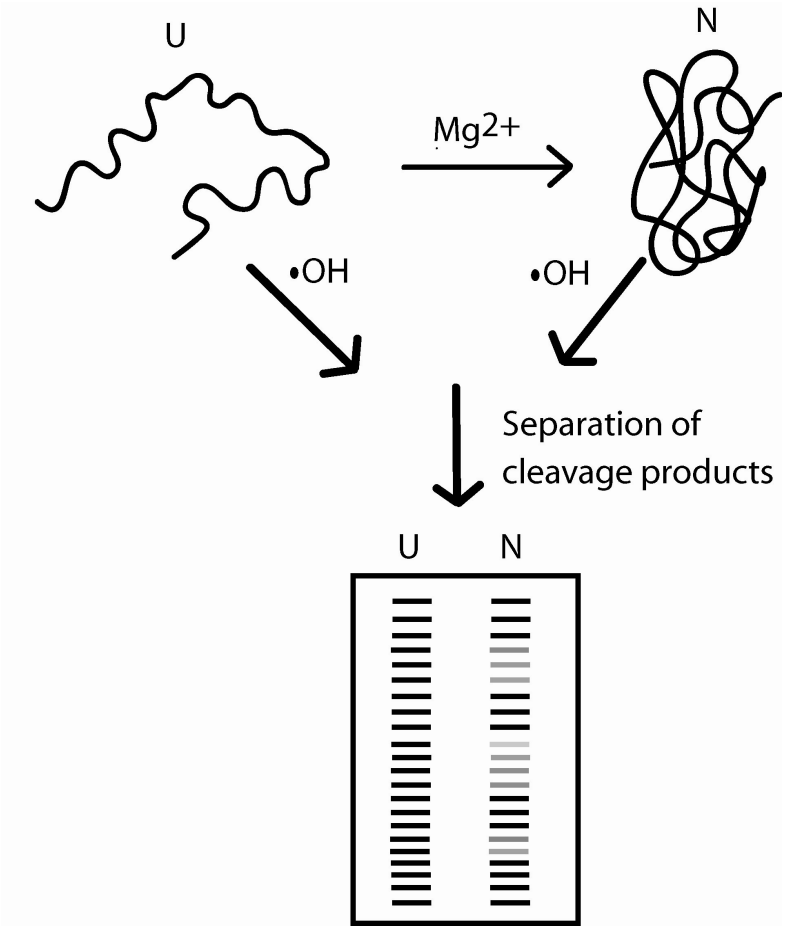

Figure 2.1: Schematic of hydroxyl radical footprinting

Unfolded RNA (U) folds to a compact native structure (N) upon the addition of Mg^{2+} Hydroxyl radicals generated, cleave the RNA backbone where it is most solvent accessible. After hydroxyl radical cleavage, reactions are quenched and cleavage products are separated on a denaturing gel. The lane labeled "U" shows cleavage products obtained for the unfolded RNA. The unfolded molecule cleaves uniformly giving rise to cleavage products at each residue. The cleavage products for "N" show regions where bands are less intense or protected. These regions are relatively inaccessible to solvent and thus cleaved less by hydroxyl radicals. Figure adapted from [84, 87].

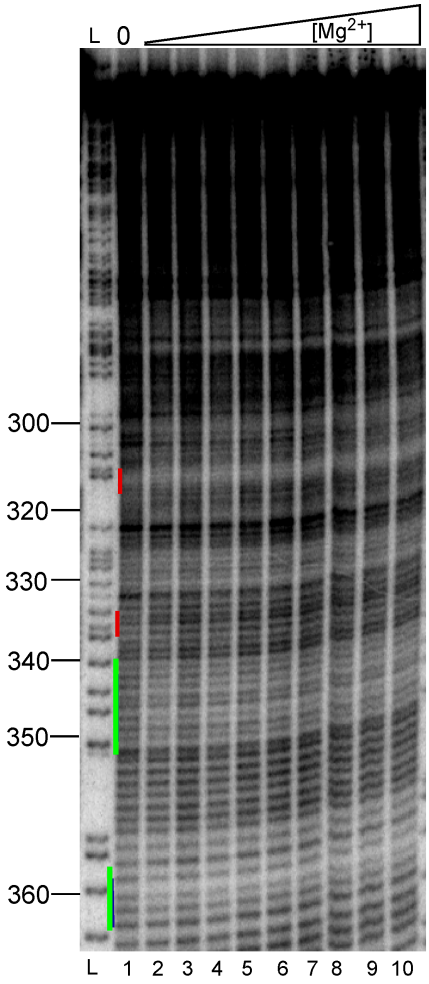

Figure 2.2: Equilibrium folding of $E^{\Delta P5abc}$ ribozyme

Fe(II)-EDTA footprinting of 3' end-labeled $E^{\Delta P5abc}$ allowed to fold in presence of 0-50 mM Mg^{2+} (lanes 1-10) to generate equilibrium mixtures. Lane L is RNA ladder of guanine nucleotides generated by RNase T1 digestion under denaturing conditions (7M urea). Green bars refer to nucleotides that become protected with increasing Mg^{2+} and red bars refer to nucleotides that become enhanced. Bands were quantified using the SAFA software.

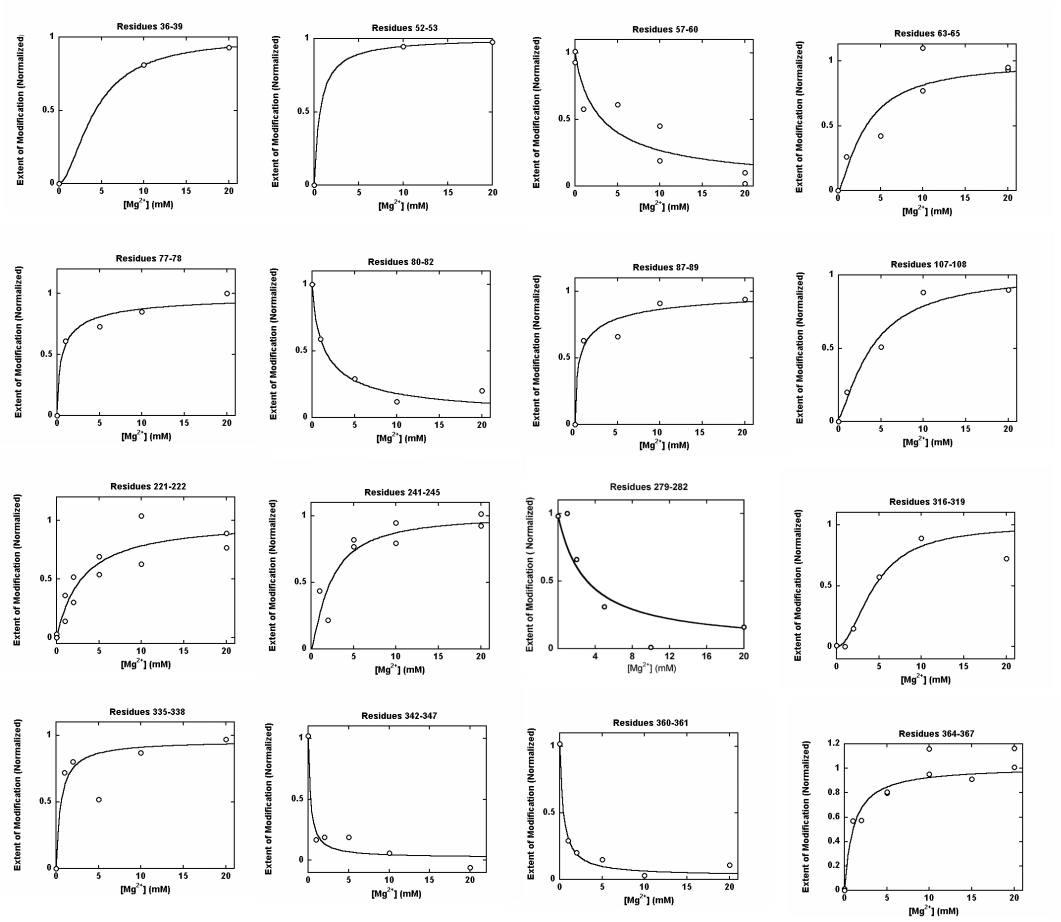

Figure 2.3: Mg^{2+} dependent folding of $E^{\Delta P5abc}$

Plots showing regions of protections and enhancements as a function of Mg^{2+} concentrations. Data were fit to Hill equation ($[Mg^{2+}]^n / (K_{1/2}{}^n + [Mg^{2+}]^n) + C$ where $K_{1/2}$ represents the magnesium requirements to give half of the native ribozyme formed, n is the Hill coefficient and C is the offset which is the fraction of ribozyme that folds directly to the native state. A summary of the $K_{1/2}$ values are shown in Table 2.1. All listed regions were probed in at least four independent experiments [58].

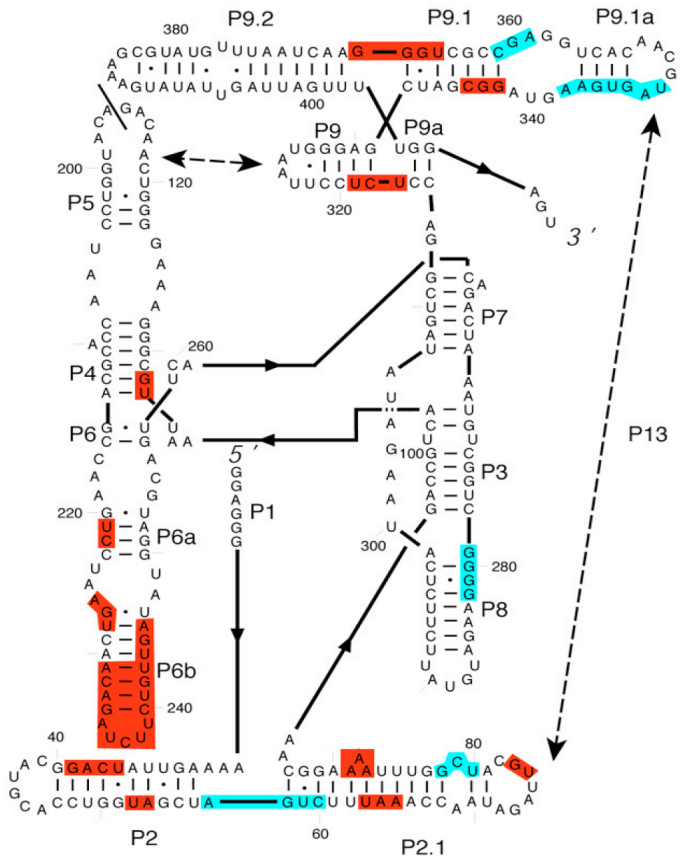

Figure 2.4: Secondary structure map of $E^{\Delta P5abc}$ highlighting changes in solvent accessibility

Regions of protections are shown in blue and enhancements are shown in red which were observed with increasing Mg^{2+}. Protections were observed in, P2 and P2.1, P8 and P9.1a. Enhancements were see in P2.1, L2.1, P2, P6b, L6b, P6a, P4, P9 and P9.1 [58].

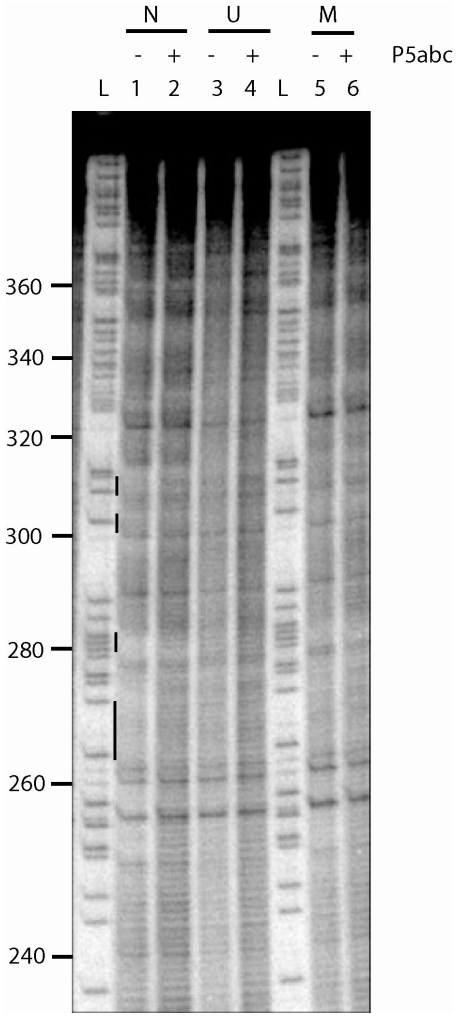

Figure 2.5: Hydroxyl radical footprinting of the native and misfolded $E^{\Delta P5abc}$

Footprinting of 3' end labeled RNA. Lane 'L' represents RNase T1 digestions of the RNA for use as G sequencing ladder. Lanes labeled 'U', 'M' and 'N'; show Fe (II)-EDTA footprinting data for unfolded, misfolded and native ribozyme respectively. Regions that gave reproducible differences between native and misfolded complex with P5abc are indicated by vertical black lines.

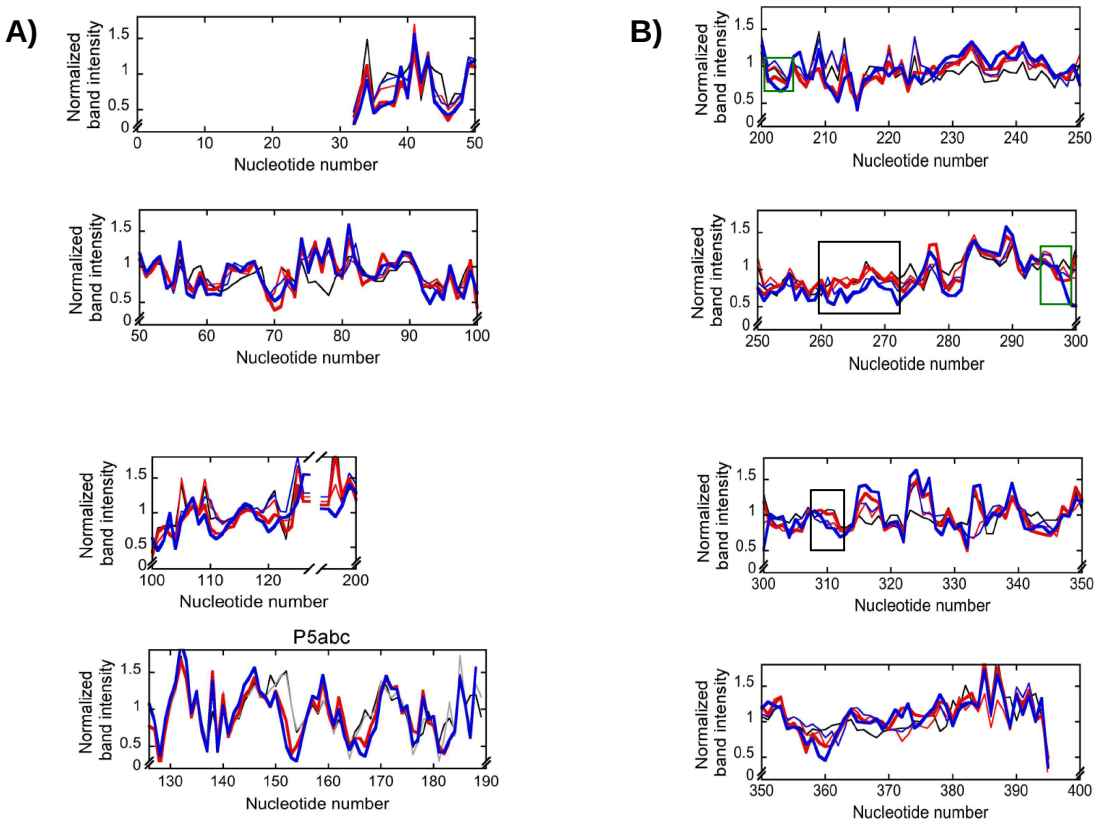

Figure 2.6: Band intensity at each nucleotide for $E^{\Delta P5abc}$ native and misfolded ribozymes

Each plot shows the normalized band intensity vs. the nucleotide number for unfolded (black line), misfolded ribozyme (thin red line), native ribozyme (thin blue line), misfolded ribozyme with P5abc (thick red line) and native ribozyme with P5abc (thick blue line). A) Plots showing the 5' region of the ribozyme color coded as explained above. Residues 126-190 show footprinting of the P5abc RNA in the absence of Mg^{2+} (black line), in the presence of Mg^{2+} (grey line) and in complex with the misfolded ribozyme (red line) and native ribozyme (blue line). B) Regions covering 3'regions of the ribozyme 200-395 color coded as described in panel A. Differences that were previously observed between the native and misfolded in the full length ribozyme are boxed in black and additional changes observed in the $E^{\Delta P5abc}$ are boxed in green [79].

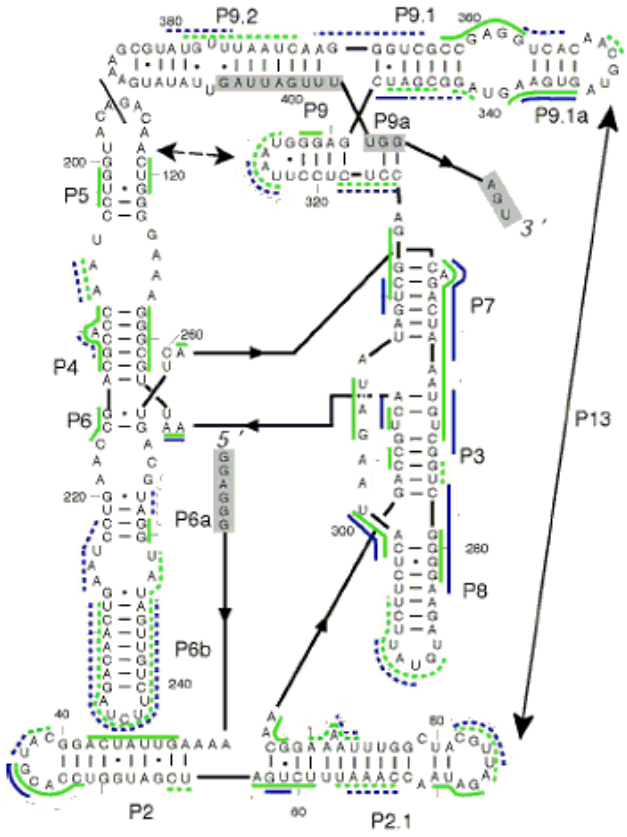

Figure 2.7: Regions of protections and enhancements for $E^{\Delta P5abc}$ alone or in the presence of P5abc

Fe (II) EDTA footprinting of $E^{\Delta P5abc}$ in the presence of P5abc (green) and in the absence of P5abc (blue). Footprinting was done under conditions to generate mostly native ribozyme. Regions that were protected in the folded structure relative to the unfolded structure (absence of Mg^{2+}) are indicated with a solid lines and curves. Regions of enhancement or increased accessibility are indicated with dashed lines or curves. The extent of protections observed in the core region when P5abc is present supports the model that P5abc compact the core helices.

Region	Observation	$K_{1/2}$, mM Mg^{2+}
36-39	Enhancement	< 5
52-53	Enhancement	< 5
57-60	Protection	3.2 ± 1.2
63-65	Enhancement	3.3 ± 1.7
77-78	Enhancement	0.8 ± 0.4
80-82	Protection	1.1 ± 0.3
87-89	Enhancement	0.7 ± 0.5
107-108	Enhancement	4.0 ± 1.5
221-222	Enhancement	3.0 ± 1.5
241-245	Enhancement	2.3 ± 1.0
279-282	Protection	3.0 ± 2.3
316-319	Enhancement	4.1 ± 1.4
335-338	Enhancement	0.5 ± 0.5
342-347	Protection	0.3 ± 0.2
360-361	Protection	0.4 ± 0.1
364-367	Enhancement	1.0 ± 0.5

Table 2.1: Summary of regions of protections and enhancement observed by Fe (II)-EDTA footprinting at varying Mg^{2+} concentration.

For each region of the ribozyme protections and enhancements were fit to a Hill equation to give the measured $K_{1/2}$ values. Standard deviation was calculated from at least four independent experiments.

Chapter 3: The tyrosyl-tRNA synthetase (CYT-18) stabilizes the native structure relative to a long-lived misfolded structure without compromising folding kinetics

3.1 INTRODUCTION

RNAs carry out a myriad of cellular functions, many of which require folding to specific structures. Thus, the sequences of these RNAs must encode functional structures that are stable enough to be populated at equilibrium and are kinetically accessible during folding [93]. Because structured RNAs typically include base-paired regions that form favorably even as isolated elements, globally unfolded conformations are generally not stable. Therefore, a critical challenge for RNA is to stabilize functional structures relative to alternative conformations that include some but not all of the native secondary and tertiary contacts and may also include non-native contacts [93, 94]. The high stability of local structure also poses challenges for the kinetics of folding, because disruptions of native or non-native contacts that are necessary to allow refolding of misfolded intermediates can be slow enough to limit the overall rate of folding [95, 96].

Since the discovery of self-splicing more than 25 years ago [97], group I intron RNAs have served as powerful models for studies of RNA folding, structure, and function [98]. Particularly valuable have been the intron from *Tetrahymena thermophila* and its 'ribozyme' derivative, which lacks exons and, instead of self-processing, cleaves an oligonucleotide substrate in a reaction that mimics the first step of splicing [51]. The structure consists of two core helical domains, P4-P6 and P3-P8, which are buttressed by peripheral domains encircling the core [99-104]. All group I introns share the core

domains, but subgroups differ in the identities and abundance of their peripheral domains [105, 106].

The peripheral domains play key roles in stabilizing the functional structures of many group I introns, including the *Tetrahymena* intron. One peripheral structure element that is particularly well-studied is P5abc. Based on comparisons of the wild-type ribozyme with a deletion variant that lacks P5abc ($E^{\Delta P5abc}$), it has been shown that P5abc greatly stabilizes the folded structure relative to partially folded intermediates and enhances catalytic activity [60, 107, 108]. P5abc also stabilizes the native state relative to a long-lived misfolded conformer that is similar in global structure and suggested to differ in topology within the core [109, 110]. This misfolded intermediate accumulates during Mg^{2+}-induced folding of both the wild-type and $E^{\Delta P5abc}$ ribozymes, [111-113] and it remains substantially populated at equilibrium in the absence of P5abc, indicating that the native and misfolded structures are equivalently stable [109]. P5abc can bind the core as a separate molecule [107], and it shifts the equilibrium to strongly favor the native state (K_{eq} = 70,000) [109].

On the other hand, the misfolded $E^{\Delta P5abc}$ ribozyme folds to the native state 100-fold faster than the wild-type ribozyme, indicating that the presence of P5abc hinders this refolding [113]. The folding transition requires extensive unfolding and is greatly accelerated by mutations that abolish native peripheral contacts, including those of P5abc [110]. P5abc forms two tertiary contacts to the P4-P6 domain, which packs against P5abc, and a third to the peripheral L2 loop, termed P14. The acceleration of refolding upon deletion of these tertiary contacts suggested that the peripheral ring can form in the context of the misfolded core, and it must be disrupted to allow the core to rearrange to

the correct topology. It has been suggested that, while peripheral elements and contacts can stabilize functional RNA structures, they may also increase the likelihood and severity of misfolded intermediates by forming contacts incorrectly or at the wrong time [113].

An alternative strategy employed by many group I introns is to stabilize functional structures by interacting with protein cofactors [114]. Some of these proteins are encoded within their cognate introns, while others are cellular proteins that have other functions and have apparently been co-opted through evolution. The *Neurospora crassa* CYT-18 protein is a prototypical member of this latter group, functioning both as a tyrosyl-tRNA synthetase and a cofactor for several mitochondrial group I introns [115, 116]. The function of CYT-18 in group I intron splicing is thought to have arisen recently in evolution, because functional intron binding depends on insertions in CYT-18 that are present only in one subphylum of fungi [117]. These insertions form a critical part of an RNA binding surface that is distinct from the tRNA binding site and allows a CYT-18 dimer to bind to the group I intron core along the face of P4-P6, in roughly the same position occupied in other introns by P5abc [72, 118-121]. Importantly, a level of functional equivalency of P5abc and CYT-18 has been demonstrated by the finding that CYT-18 gives substantial rescue of catalytic activity of the $E^{\Delta P5abc}$ *Tetrahymena* ribozyme [63].

Here, we investigate the extent to which CYT-18 replaces P5abc in folding stability and kinetics of the $E^{\Delta P5abc}$ ribozyme. We find that CYT-18 stabilizes the native state, in agreement with earlier results, and that it specifically recognizes the native state in preference to the long-lived misfolded conformation. Unlike P5abc, CYT-18 does not

slow refolding of the misfolded conformer to the native state. We suggest that this difference in behavior arises from a major difference in the nature of the interactions; whereas P5abc binding completes the ring of peripheral contacts, CYT-18 forms more extensive contacts directly with the core and its contacts with the misfolded core can apparently be maintained during refolding. The results have implications for the functions of CYT-18 in vivo and, more generally, for how proteins may evolve to specifically stabilize functional RNA structures.

3.2 MATERIALS AND METHODS

3.2.1 Preparation and labeling of RNA

Wild-type (L-21/ScaI) and $E^{\Delta P5abc}$ ribozymes and P5abc RNA were prepared by in vitro transcription using T7 RNA polymerase [51] and purified using a Qiagen RNeasy column as described previously [122]. Ribozymes were 5′ end-labeled by removing 5′-phosphoryl groups with shrimp alkaline phosphatase (MBI Fermentas, Amherst, NY) and then incubating with [γ-^{32}P]ATP and T4 polynucleotide kinase (New England Biolabs). RNA oligonucleotides (Dharmacon Research, Lafayette, CO) were 5′ end-labeled by using [γ-^{32}P]ATP and T4 polynucleotide kinase [51]. Labeled RNAs were purified by nondenaturing polyacrylamide gel electrophoresis (PAGE), using a 6% gel for ribozymes and a 20% gel for oligonucleotides.

3.2.2 CYT-18 expression and purification

CYT-18 was expressed and purified essentially as described [121, 123]. Briefly, CYT-18 was expressed from the plasmid pEX560 in the E. coli strain HMS174 (DE3). Cells were grown in 1 l of LB medium supplemented with 50 µg/ml carbenecillin to an OD_{600} of 0.6. CYT-18 expression was induced by adding IPTG (1 mM) and growing cells overnight at 25 °C. Cells were pelleted by centrifugation, resuspended in buffer A (25 mM Tris-Cl, pH 7.5, 500 mM KCl, 1 mM EDTA, 10% glycerol), and lysed by incubating with 1 mg/ml lysozyme and sonicating (5 pulses of 20 s). Cellular debris was pelleted by centrifugation (1500 × g, 30 min at 4 °C). All subsequent steps were carried out at 4 °C. Remaining nucleic acids were removed by adding polyethyleneimine (PEI) slowly from a 10% stock to 1%, followed by centrifugation. CYT-18 was precipitated by adding powdered ammonium sulfate to 40% saturation followed by centrifugation. Ammonium sulfate precipitation was repeated two additional times to remove contaminating ribonucleases. The pellet was resuspended in 30 ml of buffer A containing 500 mM KCl and dialyzed against 1 l of buffer A containing 25 mM KCl for 1 h and then overnight against an additional liter of the same buffer. CYT-18 was further purified using a heparin-sepharose column (16/10, GE Healthcare) previously equilibrated with buffer A containing 25 mM KCl. Bound CYT-18 was eluted with a linear KCl gradient (200–1000 mM), with the bulk of CYT-18 eluting in a peak centered at approximately 400 mM KCl. Fractions containing CYT-18, as determined by SDS-PAGE, were flash-frozen in liquid nitrogen and stored at –80 °C.

3.2.3 Concentration determinations of RNA and CYT-18 protein

CYT-18 concentration was determined using Bradford dye reagent (BioRad, Hercules, CA) with an IgG standard curve. This measurement gave the same values within error as absorbance at 280 nm (not shown). All CYT-18 concentrations refer to the dimer. Ribozyme concentrations were determined spectrophotometrically using the following extinction coefficients (260 nm): wild-type ribozyme, 3.9×10^6 M^{-1} cm^{-1}; and $E^{\Delta P5abc}$, 3.2×10^6 M^{-1} cm^{-1} [111].

3.2.4 Ribozyme catalytic activity assays

Ribozyme activity assays were performed essentially as described previously [109, 110, 112]. $E^{\Delta P5abc}$ ribozyme (200 nM) was first incubated with Mg^{2+} for 10 min at 25 °C to generate largely misfolded ribozyme or for 30 min at 37 °C to generate an equilibrium mixture of native and misfolded ribozyme. CYT-18 was then added and a continued incubation was performed under standard conditions (25 °C, 50 mM Na-MOPS, pH 7.0, 10 mM Mg^{2+}). At various times, the fraction of native ribozyme was determined by removing aliquots, adding trace radiolabeled substrate ($CCCUCUA_5$), quenching time points with two volumes of 90% formamide, and separating the substrate and product by 20% denaturing PAGE as described [51].

The fraction of substrate cleaved rapidly is a good measure of the fraction of native ribozyme because the native and misfolded species both bind the substrate with similar rate constants. Once bound to the native ribozyme, the substrate is rapidly cleaved, whereas substrate bound to misfolded ribozyme is not cleaved until the misfolded ribozyme refolds to the native state or the substrate is released from the

misfolded ribozyme (0.02 min^{-1}), at which point it can bind to the native ribozyme or rebind to misfolded ribozyme [112]. For these experiments, the cleavage reaction was carried out by the complex of the $E^{\Delta P5abc}$ ribozyme and CYT-18 protein, which gives a lower rate constant for cleavage than the wild-type ribozyme (~0.5 min^{-1} instead of 10 min^{-1} for the intact ribozyme). Instead of collecting only a single time point at 1 min to indicate the amplitude of the cleavage burst [109, 110, 112], we carried out complete time courses of substrate cleavage and determined the burst amplitude directly by fitting progress curves with an equation that included two exponential terms using Kaleidagraph. Results are given as the average and standard deviation of two to six replicate measurements.

3.2.5 CYT-18 binding and dissociation kinetics

CYT-18 association with the $E^{\Delta P5abc}$ ribozyme was monitored by retention of (^{32}P)-labeled ribozyme on a nitrocellulose filter. For measurements of k_{off} from misfolded ribozyme, labeled ribozyme was first incubated at 25°C in 50 mM Na-MOPS, pH 7.0, 10 mM Mg^{2+} for 10 min to generate a population of predominantly misfolded ribozyme. CYT-18 was then added to a final concentration of 200 nM and incubated for 5 min to allow binding. For measurement of k_{off} from the native ribozyme, the ribozyme was pre-incubated with CYT-18 under the same solution conditions for 30 min at 37 °C to allow the ribozyme to refold completely to the native state. Excess unlabeled ribozyme was then added (1.6 µM) and at various times, 10 µl aliquots were applied to a nitrocellulose filter that was pre-soaked in reaction buffer and backed with a nylon membrane. Filters were washed immediately with 1 ml of the same buffer and then removed and dried.

Labeled RNA on the nitrocellulose and nylon membranes was quantitated by using a phosphorimager. The fraction of labeled material was plotted over time and rate constants were determined using Kaleidagraph (Synergy Software).

Experiments to measure binding kinetics to the native ribozyme were performed similarly. Radiolabeled ribozyme was first incubated with 10 mM Mg^{2+} for 30 min at 37°C to allow equilibration between the native and misfolded species. CYT-18 was added and, at various times thereafter, aliquots were withdrawn, added to a chase of unlabeled ribozyme, and then applied to a filter and processed as described above.

3.3 RESULTS

In the sections below, a thermodynamic cycle formalism is used to analyze CYT-18 stabilization of the native structure of the $E^{\Delta P5abc}$ ribozyme relative to the long-lived misfolded structure (Figure 3.1). For the ribozyme alone, the native and misfolded conformations are approximately equal in stability ($K_{M\leftrightarrow N}^{E^{\Delta P5abc}}$ = 1.4, top of Figure 3.1) [109], and we used catalytic activity assays to show directly that CYT-18 increases the fraction of native ribozyme by stabilizing it relative to the misfolded form. From Figure 3.1 it can be seen that the magnitude of the stabilization conferred by CYT-18 is equal to the difference in its binding affinities for the native and misfolded forms. By analogy, we showed previously that the P5abc peripheral element binds 50,000-fold tighter to the native ribozyme, indicating that it stabilizes the native fold relative to the misfolded conformation by an equivalent amount, or approximately 6 kcal/mol [109]. Therefore, we used binding assays to confirm that CYT-18 binds preferentially to the native ribozyme and to probe the magnitude of the difference. In subsequent sections, we measured the

effects of CYT-18 on the kinetics of folding from the misfolded intermediate to the native state.

3.3.1 CYT-18-mediated stabilization of the native state detected by RNA catalytic activity assay

We used a ribozyme activity assay to determine whether binding of CYT-18 shifts the equilibrium of the $E^{\Delta P5abc}$ ribozyme toward the native structure. Thus, we incubated $E^{\Delta P5abc}$ to allow formation of an equilibrium mixture of the native and misfolded species [109]. Then we added CYT-18 and monitored the fraction of native ribozyme over time by measuring the fraction of trace labeled substrate (S*), added subsequently, that was cleaved in a rapid burst (Figure 3.2A) [109, 110, 112, 113]).

For each incubation time with CYT-18 we collected a full time course of the cleavage reaction to unambiguously determine the fraction of native ribozyme over time (Figure 3.2B). The substrate cleavage reaction with bound CYT-18 was slower than with P5abc (~0.5 min^{-1} compared to ~10 min^{-1}). Nevertheless, cleavage was much faster than dissociation of S* and refolding of misfolded ribozyme, such that the amplitude of the cleavage burst provided a good measure of the fraction of native ribozyme (see Materials and Methods). After addition of CYT-18, the fraction of S* that was rapidly cleaved gave a clear time-dependent increase (Figure 3.2C). By this measure, the fraction of native ribozyme reached the maximum obtained with P5abc, which indicates the absence of significant misfolded ribozyme. (Not all of the substrate is cleaved rapidly because of a small, preparation-specific fraction of ribozyme that is inactive, most likely from damage) [109].

To confirm that CYT-18 shifted the equilibrium toward the native species, we used the same assay to monitor the re-appearance of misfolded ribozyme after incubation with CYT-18 and subsequent inactivation by proteolysis (Figure 3.3A). As expected, the fraction of native ribozyme decreased again after removal of CYT-18, reaching the same endpoint as a reaction in which folding was monitored in the absence of CYT-18 (Figure 3.3B). These two reactions also gave essentially the same rate constant, as expected for a two-state equilibrium between the native and misfolded conformations.

Together, these results show that CYT-18 stabilizes the native state of the ribozyme relative to the misfolded form, indicating that it binds preferentially to the native core (see Figure 3.1). The effect is large enough that the misfolded ribozyme is undetectable. With an uncertainty in the cleavage burst amplitude of 1-2%, we estimate that CYT-18 generates an equilibrium of at least 50, a stabilization of $\geq$ 35-fold from the equilibrium value of 1.4 in its absence ($\geq$ 2.1 kcal/mol).

3.3.2 Tight and specific binding of CYT-18 to the native $E^{\Delta P5abc}$ ribozyme

To determine the equilibrium constants for CYT-18 binding to the native and misfolded ribozyme, we set out to measure the kinetics of binding and dissociation. It is not possible to measure the equilibrium constants directly because CYT-18 binds to the native core of group I introns tightly [116, 123], and at the low concentrations of CYT-18 that would be required, CYT-18 could dissociate from functional dimers into monomers [123]. Further, the native and misfolded $E^{\Delta P5abc}$ ribozyme interconvert on the same time scale as equilibrium is reached with subsaturating CYT-18 concentrations (data herein), complicating binding measurements to each conformer separately.

To measure the kinetics of dissociation, we used a pulse-chase procedure, trapping complexes of CYT-18 and radiolabeled $E^{\Delta P5abc}$ by nitrocellulose filter binding. With an initial pre-incubation to generate largely native ribozyme [109], the major phase of dissociation was easily followed (0.024 ± 0.018 min^{-1}, Figure 3.4). This process was considerably faster than dissociation of P5abc (10^{-6} min–1, ref. [109]), moderately faster than dissociation of CYT-18 from its cognate introns ($<10^{-4}$ – 10^{-3} min^{-1}, refs. [116, 123]), and about 5-fold faster than CYT-18 dissociation from the non-cognate yeast intron bI5 [124]. Nevertheless, the half-life of ~25 minutes indicates a relatively stable complex of CYT-18 and the native $E^{\Delta P5abc}$ ribozyme.

In an analogous experiment with a population of ribozyme that was incubated only briefly with Mg^{2+} to generate predominantly misfolded ribozyme, the major phase was much more rapid, reaching completion before the first time point (Figure 3.4). The observed minor phase gave a rate constant similar to that above (0.042 ± 0.034 min^{-1}), suggesting that it reflected CYT-18 dissociation from the small population of native ribozyme expected under these conditions (~10%, refs. [109, 113]). Completion of the major phase before the first time point (30 s) indicates a rate constant of ≥ 3 min^{-1} for dissociation of CYT-18 from the misfolded ribozyme. Increasing Mg^{2+} concentration to 35 mM or lowering temperature to 5 °C gave modest slowing of CYT-18 dissociation from the native ribozyme (2–3-fold, data not shown) but did not slow dissociation from the misfolded ribozyme to measurable rates, suggesting that dissociation under standard conditions is considerably faster than the lower limit of 3 min^{-1}. CYT-18 does appear to bind the misfolded ribozyme in the ground state under the conditions used, because the level of retention in the absence of chase was substantially greater than the first time

point after chase addition (Figure 3.4). This result does not reflect CYT-18 free in solution 'catching' the ribozyme after being trapped on the filter itself, because pre-spotted CYT-18 gave substantially lower retention of labeled ribozyme applied immediately thereafter (data not shown). The interaction with the misfolded ribozyme most likely reflects CYT-18 binding along the natural P4-P6 interface, because the non-specific interaction of CYT-18 with the full-length ribozyme is at least an order of magnitude weaker under similar conditions [63]. Nevertheless, it remains possible that CYT-18 binds the misfolded ribozyme at an alternative site, in which case the preference of CYT-18 for the native core rather than the misfolded core would be even larger, strengthening our conclusions.

We took advantage of the difference in dissociation kinetics to follow CYT-18 binding to the native ribozyme. We added CYT-18 to an equilibrium mixture of native and misfolded radiolabeled $E^{\Delta P5abc}$ ribozyme and incubated for various times before adding excess unlabeled ribozyme and applying the solution to a filter. Because the complex formed between CYT-18 and the misfolded ribozyme is expected to dissociate rapidly upon addition of chase, the experiment monitored CYT-18 binding to the native ribozyme. Across a limited range of CYT-18 concentration (5 – 20 nM), the binding process was observable and gave a rate constant of $(1.6 \pm 0.2) \times 10^8$ M^{-1} min^{-1} (Figure 3.5), ~10-fold lower than for CYT-18 binding to its cognate group I introns and to the yeast bI5 intron [116, 123, 124]. It was not possible to follow CYT-18 binding to the misfolded ribozyme with these methods because the rapid dissociation sets a lower limit on the observed rate constant ($k_{obs} = k_{off} + k_{on}$ [CYT-18]) that is too large to observe by hand.

3.3.3 Weaker binding of CYT-18 than expected from kinetics measurements

In the binding kinetics measurements above (Figure 3.5A), the final extents of RNA retention were much lower than expected from the equilibrium dissociation constant calculated from k_{off}/k_{on}. The calculated dissociation constant is ~ 0.2 nM, but concentrations of CYT-18 as high as 20 nM gave incomplete binding. Similar observations have been reported for CYT-18 binding to the cognate intron ND1 [116] and for the non-cognate intron bI5 [124]. The incomplete binding suggested one or more of the following: 1) CYT-18 is largely inactive for group I intron binding, such that the concentrations of active protein are much lower than the total concentrations; 2) a relatively weak complex is formed rapidly and rearranges slowly to a tighter complex [124, 125]; or 3) CYT-18 dimers dissociate into monomers at low nanomolar concentrations, generating a coupled equilibrium between dimerization and RNA binding and therefore weaker binding than expected.

To test the possibility of inactive CYT-18 in our preparations, we varied the ratio of CYT-18 to RNA to determine a stoichiometry (Figure 3.6). At each concentration, the fraction of bound ribozyme was determined by catalytic activity as above. Because one CYT-18 dimer binds and activates one ribozyme molecule [117, 123], a greater CYT-18 requirement would suggest the presence of inactive protein. With 100 nM ribozyme, we obtained a stoichiometry of approximately one RNA per CYT-18 dimer, indicating that most of the protein was active. At low ribozyme concentration (1 nM), activation was incomplete even with 10-fold excess CYT-18 (data not shown), confirming the results of Figure 3.5 that indicate incomplete binding at these concentrations.

We explored the second possibility, a slow transition from an intermediate to a fully-assembled RNP complex, by adding CYT-18 to the pre-folded, native $E^{\Delta P5abc}$ ribozyme and measuring substrate cleavage. If a conformational transition were slower than the cleavage step, the observed cleavage reaction would be slower when initiated by CYT-18 than when CYT-18 was pre-incubated with the ribozyme. In contrast to this expectation, the rate constant was the same, 0.4 min^{-1}, whether the reaction was initiated by CYT-18 addition or by addition of guanosine cofactor with CYT-18 pre-bound (Figure 3.7). Although CYT-18 binding to this RNA almost certainly involves intermediates [118, 123], formation of the functional complex apparently does not involve intermediates that persist on the experimental time scale, ruling out the possibility of an intermediate along this pathway causing the difference between the calculated value and the direct binding measurements.† (see footnote). This result leaves open the possibility of a slow transition to a tight but inactive complex, which could also cause the difference. However, such a model is ruled out by other experiments because it would predict that the fraction of active complex would decrease over time after CYT-18 addition, which is not observed (see Figure 3.9).

Thus, we are left with the possibility that binding is weakened because of dissociation of CYT-18 dimers to monomers. Simulations showed that CYT-18 dimerization in the low μM range, coupled to ribozyme binding of the dimer with the measured K_d of 0.2 nM, could account quantitatively for the binding data in Figure 3.5 and additional experiments (data not shown). To our knowledge, the CYT-18 dimerization constant has not been measured under the conditions of our experiments, but an upper limit of 70 nM was established under higher ionic strength conditions [123], and

a related bacterial synthetase dimerizes in the pM range [126]. Nevertheless, dimerization is highly sensitive to the details of the interface, as a single point mutation of the bacterial synthetase increases the dimerization constant to 30 μM [126], leaving open the possibility of monomers under the conditions of our binding experiments. Monomerization in solution would also lead to an underestimated binding rate constant, perhaps explaining why the binding rate constant we measured (Figure 3.5) is lower than previous results at higher ionic strength with other RNAs [116, 123, 124]. Thus, a pre-formed CYT-18 dimer may bind the ribozyme up to 10-fold faster than indicated by our measured value, approaching the diffusion limit for molecules of this size and complexity [127]. This uncertainty is reflected in Figure 3.1 as a range for the value of $K_{d,\text{CYT-18}}^{N}$, but does not affect the conclusions.

Taken together, results in the preceding sections indicate that CYT-18 binds the native ribozyme core in strong preference to the misfolded core (see Figure 3.1). Activity measurements show directly that CYT-18 binds at least 35-fold tighter to the native core. Further, CYT-18 dissociates at least 100-fold faster from the misfolded RNA, suggesting a larger difference than the lower limit established by activity. The affinity difference is unlikely to be more than approximately 1000-fold, because the misfolded RNA is bound by CYT-18 with a K_d of ≤100 nM (Figure 3.4). Thus, CYT-18 provides considerable specificity, stabilizing the native state by 2 – 4 kcal/mol relative to the misfolded intermediate.

3.3.4 CYT-18 promotes formation of the native ribozyme

To probe the ability of CYT-18 to favor formation of the native ribozyme in preference to folding intermediates more broadly, we used a previously-developed activity assay [113]. In the presence or absence of CYT-18, we varied the Mg^{2+} concentration across the range required for stable formation of the native state. After an extended incubation with or without CYT-18, we raised the Mg^{2+} concentration to support the ribozyme's catalytic activity and used this activity to determine the fraction of native ribozyme. In the absence of CYT-18, most of the RNA that is non-native during the low Mg^{2+} incubation misfolds when the Mg^{2+} concentration is raised, so this method measures the fraction of native ribozyme at the initial Mg^{2+} concentration [112, 113].

The presence of CYT-18 substantially decreased the Mg^{2+} concentration required for accumulation of the native ribozyme (Figure 3.8). Specifically, the $K_{1/2}$ decreased from 4 mM to 0.9 mM Mg^{2+} in the presence of 600 nM CYT-18, a concentration chosen because it is well above the K_d. The decreased Mg^{2+} requirement indicates that at the low Mg^{2+} concentrations (0.5 – 4 mM), CYT-18 either stabilizes the native structure such that it accumulates, or it stabilizes RNA folding intermediates that avoid misfolding when the Mg^{2+} concentration is increased. Previous results provide strong support for the first model. The $E^{\Delta P5abc}$ ribozyme lacks global tertiary structure at low Mg^{2+} concentrations [108, 109, 113], and CYT-18 induces footprinting patterns that resemble the native state, as well as activating self-splicing of the intron [63]. It is possible that CYT-18 also favors folding intermediates that subsequently avoid misfolding, which could contribute to group I intron folding (see Discussion). Such an interaction would apparently require greater than 0.5 mM Mg^{2+} under these conditions, because reactions with lower Mg^{2+}

concentrations gave only 10% native ribozyme upon subsequent folding, the same value as in the absence of CYT-18 (Figure 3.8).

3.3.5 CYT-18 does not slow refolding of the misfolded ribozyme

Previous work has shown that P5abc has a large inhibitory effect on the kinetics of refolding from the misfolded conformation to the native state, as the wild type ribozyme refolds ~80-fold slower than the $E^{\Delta P5abc}$ ribozyme [113]. To determine whether CYT-18 also slows refolding, we added various concentrations of CYT-18 to the misfolded $E^{\Delta P5abc}$ ribozyme and used the ribozyme activity assay as above to follow native state accumulation. Strikingly, the rate of refolding was not changed by CYT-18 at concentrations up to at least 2.2 μM (Figure 3.9A & B). The endpoint increased in the presence of CYT-18, as expected, because CYT-18 shifted the equilibrium toward the native state. Although it did not affect the refolding kinetics, CYT-18 apparently binds to the misfolded ribozyme at these concentrations (see Figure 3.4). If it were required to dissociate during refolding, CYT-18 would inhibit refolding at concentrations that support binding in the ground state. Thus, these results suggest that CYT-18 can remain bound without interfering with the refolding process.

An obvious difference is that P5abc is covalently linked to the ribozyme but CYT-18 is not. To ensure that inhibition by P5abc did not require covalent attachment, we measured refolding of $E^{\Delta P5abc}$ with P5abc added as a separate molecule under the same conditions used for CYT-18. As expected, P5abc strongly inhibited refolding even when added in *trans* (Figure 3.9B).

3.4 DISCUSSION

Using the $E^{\Delta P5abc}$ variant of the *Tetrahymena* ribozyme, we found that CYT-18 can discriminate between the native group I intron core and a misfolded form that is globally similar and of comparable stability. Ribozyme activity assays showed directly that CYT-18 binds more tightly to the native core than the misfolded core, with a difference of at least 35-fold, and binding experiments suggested a larger difference (100–1000-fold). Thus, by binding tighter to the native form, CYT-18 stabilizes the native structure relative to the misfolded structure. Interestingly, this stabilization is not accompanied by a decreased rate for the folding transition from the misfolded to the native conformation, even under conditions that allow CYT-18 binding to the misfolded conformation. Apparently CYT-18 binds no more tightly to the misfolded intermediate than it does to the transition state ensemble that separates the native and misfolded conformers. This behavior contrasts with that of P5abc, which greatly stabilizes the native state but inhibits the kinetics of the transition because it also binds and stabilizes the misfolded RNA. As described below, this difference, in the context of structural considerations, suggests important differences in the mechanisms of CYT-18 and P5abc in distinguishing the native and misfolded group I intron cores.

3.4.1 Interactions of CYT-18 with the conserved core of group I introns

CYT-18 interacts extensively with the P4-P6 domain of the group I intron core, with contacts extending from P5a, at the ‘top’ of the P4-P6 stack, down to J6/6a toward the bottom of the core domain [72, 101, 128, 129]. These contacts crudely resemble those of the RNA peripheral element P5abc [72]. However, there are important differences. In

addition to contacts with the rest of the P4-P6 domain, P5abc contacts the loop of P2 to form the peripheral contact P14, which is necessary to generate a ring of peripheral elements and connections that encircle the core [100]. Formation of this ring contributes strongly to stabilization of the core and is also critical for stabilization of the native state relative to the long-lived misfolded conformation [109]. Although the ring of contacts can also form around the misfolded core [110], it is much more favorable in the context of the native core, as deletion of P5abc or other ring contacts destabilizes the native state relative to the misfolded state (ref. [109] and T. Johnson and R.R., unpublished). In contrast, the group I introns that are dependent on CYT-18 either do not include a P2 element or use P2 to form a tetraloop-receptor interaction with P8, presumably making it unavailable to interact with CYT-18. Thus, CYT-18 is likely to use distinct strategies from P5abc to stabilize the native core structure relative to unfolded and misfolded conformations.

But how is this stabilization achieved? The misfolded Tetrahymena ribozyme is globally similar to the native ribozyme. In the core, it possesses all of the native secondary structure but is more exposed near the junction of the P4-P6 and P3-P8 domains and on the surface of P7 [110]. It is suggested to differ in topology from the native structure, with two strands crossing incorrectly within the core. Although there are several possibilities for the identities of the strands that are inappropriately crossed, a likely model is the incorrect crossing is between the strands J3/4 and J6/7, which connect the P4-P6 and P3-P8 domains. Uncrossing and recrossing of these strands is expected to require extensive unfolding of the periphery, accounting for the slow exchange and strong dependence on peripheral tertiary contacts.

The recent co-crystal structure of CYT-18 bound to the Twort group I ribozyme suggests mechanisms by which CYT-18 may specify the native conformation without relying on peripheral contacts (Figure 3.10) [72]. Notably, an N-terminal helix that is specific to CYT-18 and its close relatives (helix H0) inserts directly into the core and forms intimate contacts with J3/4, which may enforce the correct geometry in that region. Supporting this interpretation, CYT-18 has been shown previously to rescue the effects of mutations that weaken the triple helix interactions of J3/4 with the P4-P6 domain and to depend strongly on nucleotide identity within J3/4 for optimal binding [119, 130-132].CYT-18 also contacts P4, which forms triple helix interactions with J6/7 in the same region. Together, these interactions may be critical for the ability of CYT-18 to distinguish between the native and misfolded forms of the core. Intriguingly, deletion of helix H0 gives a CYT-18 variant that promotes splicing of a cognate intron with wild-type efficiency but with a decrease in the fraction that splices efficiently [133]. It is possible that the variant is less able to distinguish the native group I intron core from a misfolded core but is still capable of promoting efficient splicing of the fraction present in the native form.

Other interactions of CYT-18 are certain to be important for stabilizing folded structures and may also stabilize the native structure relative to the misfolded structure. Another region specific to CYT-18, Ins1, contacts the major groove of P6 and P6a and may facilitate the interaction of J6/6a with the P3 domain. Because nucleotides in the vicinity of this interaction display reduced protection in the misfolded RNA [110], this contact with CYT-18 is likely to contribute to specificity. CYT-18 also interacts with the minor groove of P9 to stabilize the tetraloop-receptor interaction of L9 and P5. Although

the L9-P5 contact apparently forms in the misfolded conformation [110], it is possible that the specific molecular arrangements stabilized by CYT-18 favor the native conformation. Last, it is possible that specificity is contributed by the C-terminal domain of CYT-18, which is present in our experiments but absent in the crystal structure. This domain is not necessary for CYT-18-enhanced splicing of the $E^{\Delta P5abc}$ intron [133], but it increases the affinity for the intron (G. Mohr and A. Lambowitz, personal communication) and could also contribute to the binding specificity observed here.

It should be noted that the specificity generated by P5abc is larger than for CYT-18, such that a model could be envisioned in which P5abc forms the same contacts as CYT-18 with the ribozyme core and generates just as much specificity from them, but P5abc further enhances specificity by forming the peripheral interactions. However, this model does not appear to hold. Contacts with J3/4 and J6/7 are notably absent for P5abc, and although a contact between an A-rich bulge in P5a and the P4-P6 stack specifically stabilizes the native state [134], this interaction is removed from detected differences between the native and misfolded structures and most likely functions as a critical component of the ring of peripheral contacts [110]. The importance of the periphery for P5abc is further underscored by the finding that disrupting P14 weakens P5abc binding to the native $E^{\Delta P5abc}$ by 108-fold (A.C., D. Kung, and R.R., unpublished results), 1000-fold larger than the full effect of P5abc in specifically binding and stabilizing the native state [109].

The importance of CYT-18 connections with the core rather than the periphery may also underlie the ability of CYT-18 to perform this stabilization without slowing refolding of the misfolded conformation. Although this transition is more facile than for

the full-length ribozyme, it is strongly accelerated by urea, indicating substantial unfolding [113]. The strong association of CYT-18 with the P4-P6 core domain rather than with the periphery may allow unfolding of the periphery, and possibly even the P3-P8 domain, without disruption of contacts with CYT-18.

3.4.2 Functional interactions of CYT-18 with its cognate group I introns in vivo

It is striking that CYT-18 stabilizes the native form of the Tetrahymena group I intron relative to the globally similar misfolded conformer, despite not encountering this RNA in vivo. It is possible that CYT-18 functions in part to stabilize the native forms of its cognate introns relative to analogous misfolded forms. Because all group I introns share the same core structure and topology, they likely have the potential to form the same misfolded structure. Further, at least two of the three group I introns that interact with CYT-18 misfold in vitro (ref. [135] and D. Mitchell and R.R., unpublished results), although structural features of the misfolded structures are unknown.

In vivo, CYT-18 is likely to encounter its cognate group I introns initially as folding intermediates, because they do not form stable tertiary structure in its absence. Prior work has suggested that CYT-18 first forms contacts with the P4-P6 domain and then induces or captures correct structure in the P3-P8 domain [118, 136]. In the context of this model, when CYT-18 contacts the P4-P6 domain it may also stabilize connections to the P3-P8 domain, enforcing the correct topology at this region and formation of base triples between the domains [72]. Formation of native structure and connections could bias folding to the native state, which could then be captured by additional interactions with P3-P8. In this regard, it is interesting to consider that some of the increased native

ribozyme observed after incubation at low Mg^{2+} concentrations (Figure 3.8) may reflect CYT-18-stabilized folding intermediates that avoid misfolding upon subsequent addition of higher Mg^{2+} concentration.

By interacting with intermediates to favor pathways that avoid misfolding, CYT-18 would function as a chaperone in addition to stabilizing the functional structure. However, the three introns that bind CYT-18 in vivo also depend on the DEAD-box protein CYT-19 for splicing, indicating that any chaperone activity of CYT-18 is not sufficient to avoid misfolding altogether. It is possible that some non-native structure includes exons [137, 138], which would most likely not be affected by CYT-18, or involves large changes in secondary and/or tertiary structure that render the intron unrecognizable to CYT-18, such that a general RNA chaperone is required.

3.4.3 Evolution of protein co-factors in stabilization and folding of structured RNA

Analysis of the sequences and properties of mitochondrial tyrosyl tRNA synthetases from various fungal species indicated that only CYT-18 and its close relatives participate in group I intron splicing [117]. Thus, this activity apparently arose recently in evolution, presumably driven by accumulating structural defects in group I introns that had inserted into essential mtDNA sequences. As CYT-18 evolved increased efficiency in promoting splicing, the group I introns that could interact productively probably continued to lose their inherent abilities to form functional structures, becoming increasingly dependent on CYT-18.

With this evolutionary progression in mind, it is interesting to consider the activities that could be provided by a protein as it evolves to function in complex with a

structured RNA. At the earliest stages, the protein is likely to recognize only a limited structural feature of the RNA. An early ancestor of CYT-18 most likely possessed limited binding potential for the P4-P6 domain. If the fortuitous interaction is formed by a region of the protein that is not highly constrained by the existing function of the protein, further evolution can rapidly enhance binding to the RNA. Through time, the precision with which the RNA is recognized could be increased, allowing stronger binding to the correctly folded RNA than to misfolded conformations. Discrimination against misfolded conformations with large structural differences could arise solely as a consequence of evolving a surface that binds specifically to the native state, whereas discrimination against closely-related misfolded forms, such as the core misfold studied here, could require explicit selective pressure against tight binding. Discrimination against misfolded RNA species may be common, as two yeast proteins have recently been shown to stabilize the native structure of the bI3 group I intron relative to a misfolded conformation that accumulates in the absence of the proteins [139].

As a protein enhances its RNA binding potential through evolution, it is also likely to interact with folding intermediates, and these interactions can either help or hinder folding of the RNA. The ability to recognize native RNA structure may allow the protein to recognize native-like structure in folding intermediates and therefore to guide folding toward the native state. Along these lines, the yeast Cbp2 protein binds weakly to folding intermediates of the group I intron, bI5, and then traps native tertiary structure that forms transiently [140], although the early association has been suggested to enhance conformational sampling rather than restricting it. On the other hand, protein association can be a hindrance if it traps native structure that must be disrupted to allow resolution of

non-native structure elsewhere in the RNA. Although this is a potential liability for proteins, from the perspective of a structured RNA the same concern applies to RNA structural elements. Indeed, the situation may be worse for peripheral elements because they are covalently connected and because RNA-RNA contacts can be inherently long-lived. Thus, recruitment of cofactor proteins may allow RNAs to minimize peripheral elements through evolution, as may have occurred during evolution of CYT-18 and its cognate introns. It will be interesting to explore further whether CYT-18 and other specific RNA-binding proteins commonly function not only by binding functional structures, but also by providing specificity and enhancing the folding processes of their cognate RNAs.

3.5 FOOTNOTE

†Two conditions must be met for an intermediate to generate a difference between equilibrium constants determined from kinetics and equilibrium binding experiments [109]. First, the intermediate must be sufficiently long-lived to be trapped experimentally in the binding kinetics experiments. With our methodology, the intermediate would have to persist between the addition of chase and application to a filter, approximately 30 s. Second, the intermediate must partition favorably back to dissociation rather than proceeding on to form the final complex. If both conditions are met, the ratio of the rate constants governing these alternative fates will not contribute to the equilibrium constant calculated from kinetics. This is because the binding kinetics experiment will simply measure formation of the intermediate, and the dissociation experiment will not include the ratio of these rate constants because, once formed, the intermediate will go on to dissociate most of the time. However, this ratio will contribute to the true equilibrium binding expression and be present in the direct binding experiment. Because the ratio is <<1, this scenario will give artificially tight binding in kinetics experiments.

Here, our finding that the active complex is formed with an overall rate constant of at least 0.4 min^{-1} puts an upper limit of 1.7 min on the lifetime of any intermediate. This limit alone does not rule out the model, because an intermediate with this lifetime could be trapped experimentally. However, because dissociation of the intermediate must be faster than its forward progression, the lifetime of the intermediate will be decreased substantially by this dissociation. The difference between the equilibrium constant from kinetics and equilibrium measurements is ~100-fold, 0.2 nM vs 20–50 nM, so the dissociation rate constant would have to be ~40 min^{-1}. This would give any intermediate a lifetime of less than 1 s, which would not allow it to be trapped experimentally.

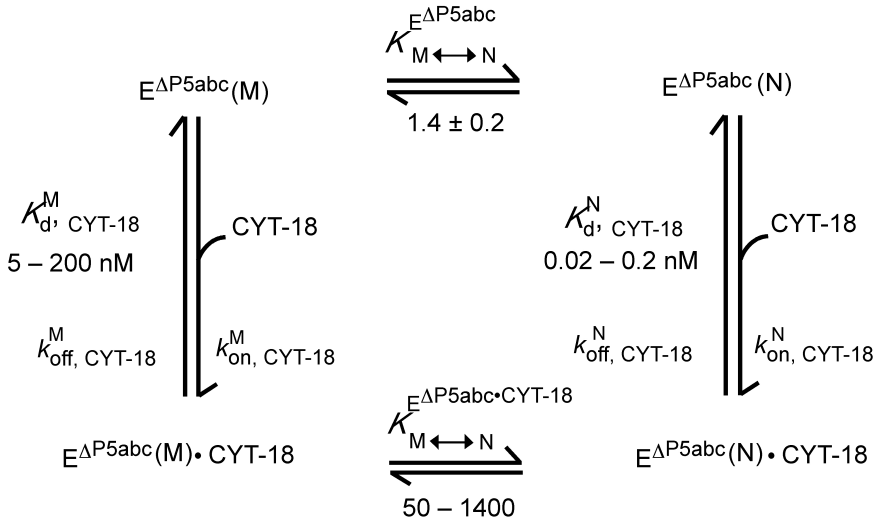

Figure 3.1: Thermodynamic cycle depicting binding of CYT-18 to native (N) and misfolded (M) $E^{\Delta P5abc}$ ribozyme

CYT-18 binding results in stabilizing of the native state. The value of 1.4 for the equilibrium between the native and misfolded forms ($K^{E^{\Delta P5abc}}_{M \leftrightarrow N}$, top of cycle) was determined previously [109].The value of $K^{N}_{d,\ CYT\text{-}18}$ was calculated from kinetics measurements to be 0.2 nM. The range shown arises from the possibility that monomerization at low CYT-18 concentrations led to an underestimated rate constant. Activity and binding measurements show that CYT-18 binds 35-1000-fold weaker to the misfolded ribozyme, leading to the range shown for $K^{M}_{d,\ CYT\text{-}18}$ (calculated from the measured value of $K^{N}_{d,\ CYT\text{-}18}$ for simplicity). The difference between CYT-18 binding affinity to the native and misfolded forms of the ribozyme leads to the calculated range of 50 – 1400 for the equilibrium between the native and misfolded conformations with CYT-18 bound ($K^{E^{\Delta P5abc \bullet CYT\text{-}18}}_{M \leftrightarrow N}$, bottom of cycle).

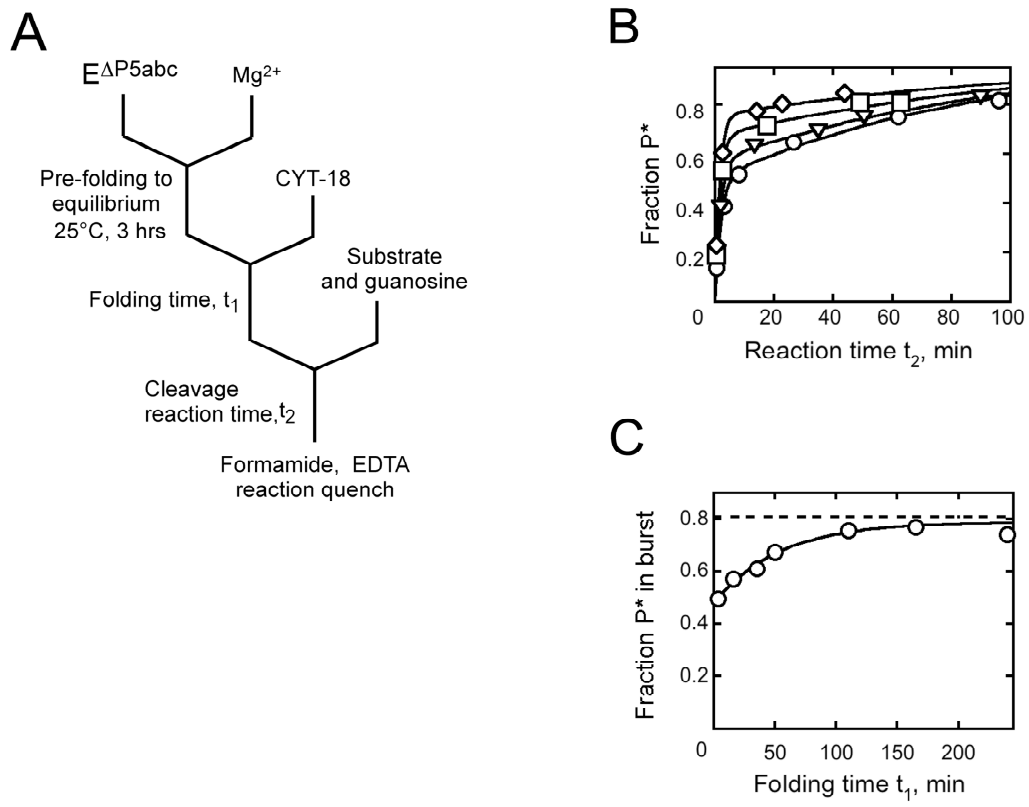

Figure 3.2: Native state stabilization by CYT-18 measured by catalytic activity of the ribozyme

A, Schematic of reaction procedure. **B,** Substrate cleavage reactions to give oligonucleotide product (P*) after incubation with CYT-18 for: 3.6 min (circles), 16 min (triangles), 110 min (squares), and 210 min (diamonds). C, The fraction of substrate cleaved rapidly (bursts from panel B) plotted against CYT-18 incubation time. Immediately after CYT-18 addition, the fraction of substrate cleaved was 0.48 ± 0.02. The fraction of misfolded ribozyme is inferred from the difference between this value and the maximum burst amplitude after extended incubations with P5abc at elevated temperature (0.81 ± 0.02, dashed line). The ratio of native to misfolded ribozyme gives an equilibrium constant of 1.5 ± 0.1 for folding of $E^{\Delta P5abc}$ ribozyme, in good agreement with the value of 1.4 reported earlier [109]. At later times, the fraction of native ribozyme increased, with k_{obs} = 0.022 ± 0.004 min^{-1}, to a value indistinguishable from that obtained by incubation with P5abc, indicating that CYT-18 shifts the equilibrium from 1.5 to at least 50.

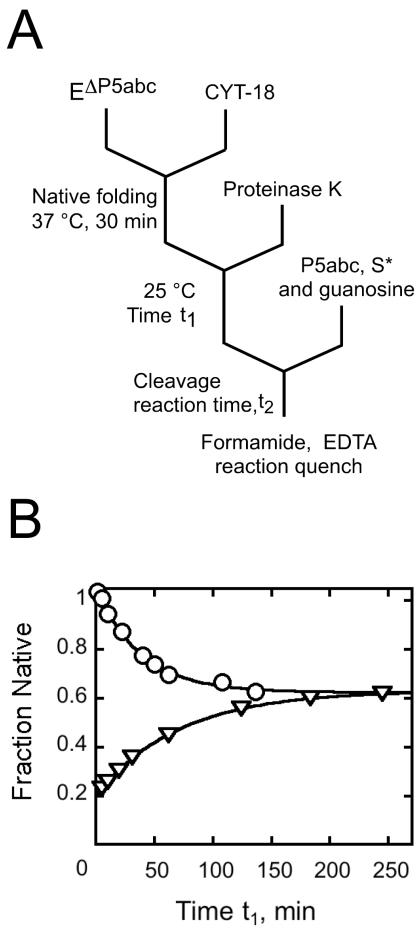

Figure 3.3: Return to equilibrium after proteolysis of CYT-18

A, reaction schematic. After allowing formation of native ribozyme, Proteinase K was added (1 mg/ml) to inactivate CYT-18. At various times thereafter (t_1), the fraction of native ribozyme was determined by activity. P5abc was added to activate the ribozyme for substrate cleavage without allowing significant redistribution of the native and misfolded forms. **B,** Circles, the fraction of native ribozyme decreased from a value reflecting essentially 100% native ribozyme (0.86) to a value reflecting the expected equilibrium between the native and misfolded species (0.51 ± 0.01, an equilibrium constant of 1.5). The decrease gave a rate constant of 0.024 ± 0.003 min^{-1}. A control reaction was performed in which the $E^{\Delta P5abc}$ ribozyme was incubated with $Mg^{2+,}$ and the approach to equilibrium was monitored. This reaction (triangles) gave an identical endpoint (0.52 ± 0.01) and rate constant (0.022 ± 0.005 min^{-1}).

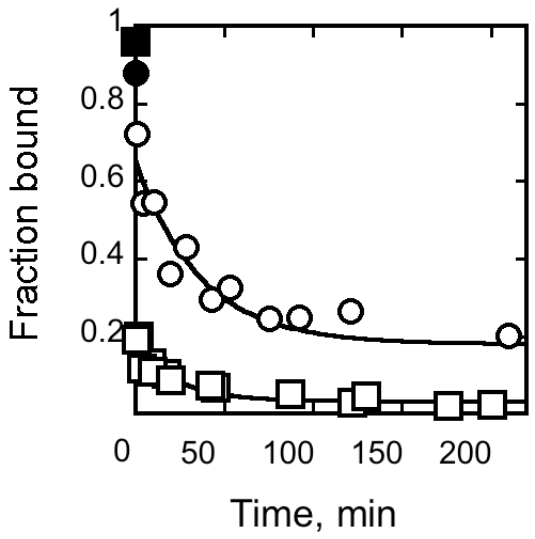

Figure 3.4: CYT-18 dissociation kinetics

CYT-18 was incubated with radiolabeled $E^{\Delta P5abc}$ ribozyme to give populations of largely native or misfolded ribozyme (circles and squares, respectively; see Methods). Complex dissociation was then followed by adding excess unlabeled ribozyme. Filled symbols show results from aliquots spotted without adding unlabeled ribozyme (again, circles and squares represent native and misfolded ribozyme, respectively). For the reaction with native ribozyme, the major phase gave a rate constant of 0.028 ± 0.010 min^{-1}. The endpoint, 0.18 ± 0.04, is the value expected from the 5-fold excess unlabeled ribozyme relative to CYT-18. For the misfolded ribozyme, most of the labeled RNA dissociated before the first time point. The observed minor phase gave a rate constant of 0.042 ± 0.013 with an amplitude of 0.11. The endpoint, 0.03, is lower than for the native ribozyme, despite the same 5-fold excess unlabeled ribozyme. This result is expected because the labeled misfolded ribozyme is expected to compete poorly with the unlabeled ribozyme, which includes a large fraction of native ribozyme.

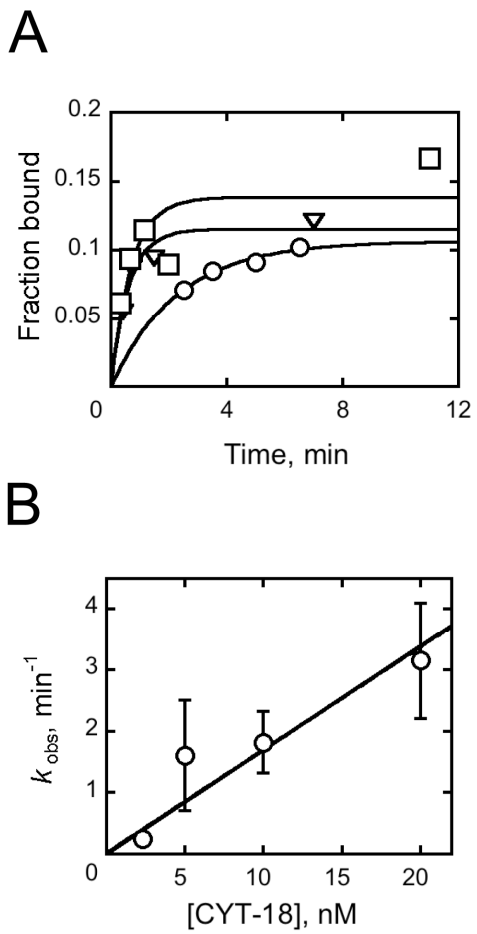

Figure 3.5: Kinetics of ribozyme binding by CYT-18

A, Time courses of binding for CYT-18 concentrations of 5 nM (circles), 10 nM (triangles), and 20 nM (squares). Aliquots were quenched with unlabeled ribozyme and then applied to filters. **B,** Rate constant for CYT-18 binding plotted against CYT-18 concentration. The slope gave a k_{on} value of $(1.6 \pm 0.2) \times 10^8$ M^{-1} min^{-1}. Error bars show standard deviations from two to three replicate measurements.

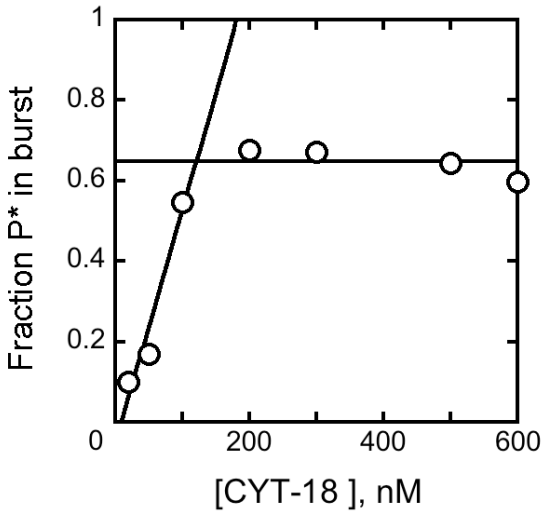

Figure 3.6: Titration of ribozyme with CYT-18 and $E^{\Delta P5abc}$ ribozyme

100 nM was pre-folded and bound to trace substrate, and various concentrations of CYT-18 were added. After 15 min of incubation, the fraction of active ribozyme was determined by measuring the fraction of S cleaved rapidly upon addition of guanosine. Full activity was observed with ~100 nM CYT-18, as indicated by the intersection of the rising and flat portions of the concentration dependence. The concentration of native, active ribozyme, as indicated by the plateau value and expected from the incubation conditions, was approximately 65 nM.

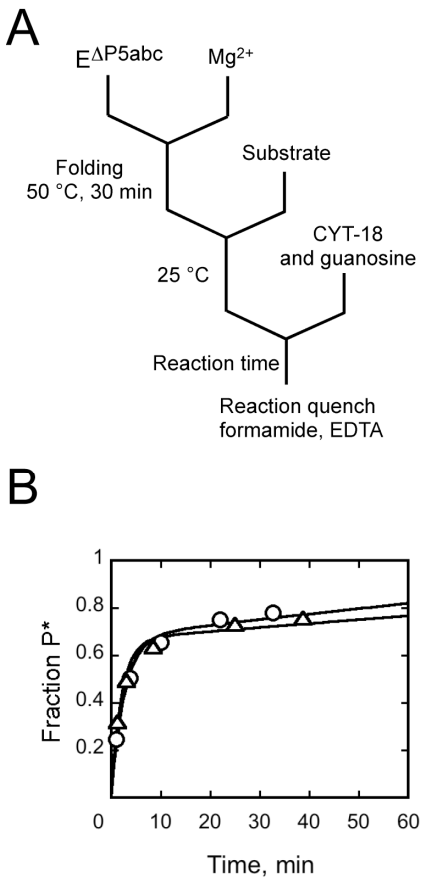

Figure 3.7: Slow conformational changes are not required upon CYT-18 binding

A, Reaction schematic. Cleavage reactions were initiated either by adding CYT-18 to pre-folded ribozyme with bound substrate, as shown, or by adding guanosine after a 15 min pre-incubation with CYT-18. **B,** Progress curves for reactions initiated by addition of CYT-18 (triangles) or guanosine (circles). The progress curves were the same within error for the two reactions, with rate constants of 0.47 ± 0.07 min^{-1} and 0.37 ± 0.06 min^{-1} respectively, and amplitudes of 0.67 for both reactions.

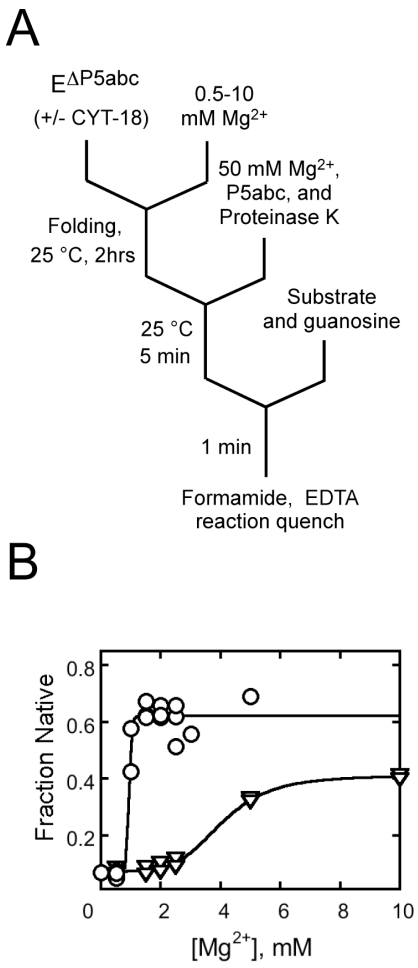

Figure 3.8: CYT-18 stabilizes the native state relative to less structured folding intermediates

A, reaction schematic. The $E^{\Delta P5abc}$ ribozyme was incubated at various Mg^{2+} concentrations to achieve equilibrium between folded and unfolded forms, and then additional Mg^{2+} was added to trap folding intermediates that subsequently misfolded and to allow determination of the fraction of native ribozyme by activity. **B,** Incubations in the absence (triangles) or presence (circles) of 0.6 μM CYT-18. Data were fit by an equation describing a cooperative transition, giving $K_{1/2}$ values of 4 mM Mg^{2+} in the absence of CYT-18 and 0.9 mM Mg^{2+} in the presence of CYT-18. The increased endpoint in the presence of CYT-18 reflects stabilization of the native ribozyme relative to the misfolded conformation.

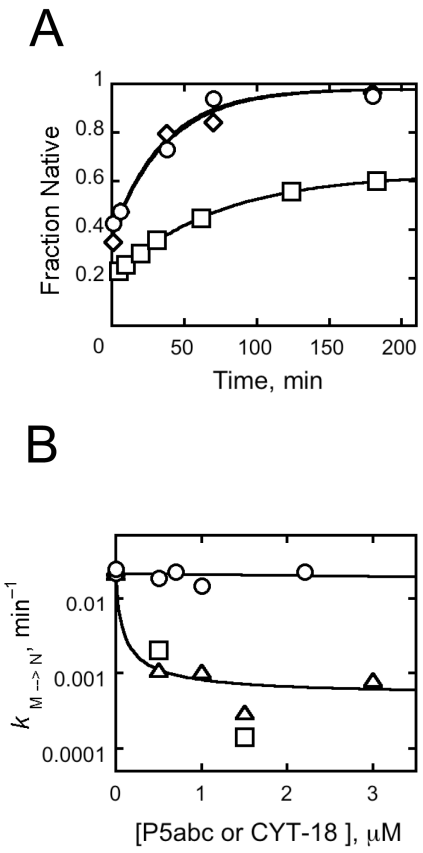

Figure 3.9: CYT-18 does not slow refolding of misfolded ribozyme.

A, Time courses of native ribozyme accumulation. After a brief incubation of $E^{\Delta P5abc}$ with 10 mM Mg^{2+} to allow misfolding, CYT-18 was added (0.7 μM, circles or 2.2 μM, triangles) and the fraction of native ribozyme was determined by activity at various times thereafter. Both reactions gave rate constants of 0.023 min^{-1}. The squares show an equivalent reaction in the absence of CYT-18, which gave a rate constant of 0.015 min^{-1}. **B,** Refolding rate constants from panel A and additional experiments are plotted against CYT-18 concentration (circles). Results from analogous reactions in which P5abc RNA was added instead of CYT-18 are also shown. Squares show reactions under identical conditions as those with CYT-18 (50 mM Na-MOPS, pH 7.0, 10 mM $MgCl_2$, plus 2.5 mM Tris-Cl, 1% glycerol, and 10 mM KCl), and triangles show reactions under standard RNA folding conditions (50 mM Na-MOPS, pH 7.0, 10 mM Mg^{2+}).

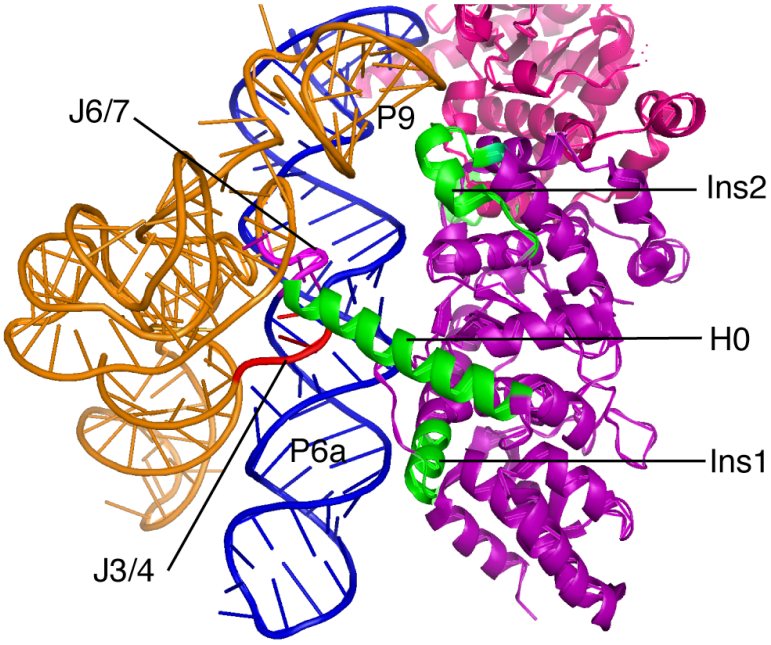

Figure 3.10: Specific contacts of CYT-18 with the ribozyme core

The P4-P6 domain is blue and the P3-P8 domain is gold. The linker regions that connect the two domains, J3/4 and J6/7, are shown in red and purple, respectively. One subunit of the CYT-18 dimer is shown in purple, with inserted regions shown in green. The other CYT-18 subunit is magenta. Coordinates are from Paukstelis *et. al*, 2008

Chapter 4: Specificity of CYT-18 to the native ribozyme relative to the misfolded ribozyme does not depend on a complete ring of peripheral contacts

4.1 INTRODUCTION

A critical aspect of understanding RNA folding is to explore how RNAs are able to populate a fully functional structure relative to all other possible folded or partially folded conformations. Structured RNAs, such as group I introns, specify the functional structure by using peripheral elements. The P5abc peripheral element from the *Tetrahymena* ribozyme confers structural specificity, stabilizing the native state by 6 kcal/mol relative to a defined misfolded intermediate [58]. Structural specificity in RNAs can also be generated by an RNA binding protein, CYT-18 , that binds the conserved core of several group I introns [21, 63, 141] and promotes the formation of the catalytically active structure [63]. In chapter 3, CYT-18 was shown to bind and preferentially stabilize the native state of the *Tetrahymena* $E^{\Delta P5abc}$ ribozyme relative to the misfolded structure, thus extending the line of evidence for structural specificity in structured RNAs.

The physical mechanisms by which P5abc and CYT-18 are able to preferentially stabilize the native ribozyme relative to a globally similar misfolded ribozyme are currently not well understood. For P5abc, a plausible model that accounts for its ability to stabilize the native state relative to the misfolded ribozyme stems from long range tertiary contacts P13, P14 and L9/P5 that encircle the core. The misfolded ribozyme has a slightly expanded core [8, 35, 60], and so cooperative formation of this circular network of long

range contacts could indeed play a role in generating specificity towards the native state. Though all the native tertiary contacts are formed in the misfolded ribozyme, the slightly bigger core may be strained. In fact, ablating the L9/P5 or P13 tertiary contacts provided evidence that the specificity is critically dependent on the formation of these of peripheral contacts surrounding the core (T. Johnson; unpublished).

The observation that a complete ring of contacts is necessary for P5abc to confer native state specificity, led us to question to what extent, if any, the peripheral ring mattered for CYT-18. It is likely however, that peripheral contacts would matter less for CYT-18 because peripheral elements are generally not well conserved and the nature of RNA/RNA contacts differ from RNA/protein contacts on a molecular level. On the other hand, both P5abc and CYT-18 bind along the same region of the P4-P6 domain and both display structural specificity to the native state. Thus, it would be interesting to test the extent to which CYT-18, like its molecular mimic P5abc, depends on a complete network of peripheral contacts.

To explore the physical origin of specificity, the role of the long range contact P14, formed between L5c and L2, was probed. Two $E^{\Delta P5abc}$ mutants were designed where the L2 sequence and the P2 stem were mutated to disrupt any potential formation of P14 when P5abc was added in *trans.* Interestingly, both mutants had similar affinity for CYT-18 as wt $E^{\Delta P5abc}$ and the differences between binding to the native state and misfolded remained large, suggesting that specificity to the native state was maintained. However, the affinity of P5abc was dramatically weakened. Ribozyme activity assay revealed that P5abc bound 10^8 fold weaker to the native state of both mutants when compared to P5abc

binding to the native wt $E^{\Delta P5abc}$ ribozyme, indicating that for P5abc, P14 is critical for native stability.

4.2 MATERIALS AND METHODS

4.2.1 Preparation of mutant ribozymes

DNA encoding changes to $E^{\Delta P5abc}$ L2 loop, P9.1 loop and ΔP2 were made by site directed mutagenesis from $E^{\Delta P5abc}$ template. Plasmid DNA was linearized using *Sca I* and RNA was made by *in vitro* transcription under the same conditions as explained in Chapter 3. Ribozyme concentrations were determined spectrophotometrically (Abs =260 nM) using the following extinction coefficient: L2 and P9.1= 3.2×10^6 M^{-1} cm^{-1} and ΔP2 stem = 3.0×10^6 M^{-1} cm^{-1}

4.2.2 Ribozyme activity assays to monitor folding.

To probe folding, the cleavage rate of these mutants must be faster than both the dissociation of the substrate from the misfolded ribozyme (0.02 min^{-1}) and overall rate of refolding. Mutant ribozymes were incubated with saturating concentrations of P5abc and CYT-18 in 50 mM Na-MOPS, pH 7, 10 mM Mg^{2+}. Cleavage reactions were done at 0.6-1 mM G with either 10 mM Mg^{2+} or 50 mM Mg^{2+}. Ribozyme activity assay to monitor folding and refolding in the presence of CYT-18 were performed as described in chapter 3.

4.2.3 Measuring P5abc binding to mutant ribozymes by activity assays.

200 nM of each mutant ribozyme was incubated with varying concentrations of P5abc (0.3–10 μM) at 25°C for 180 minutes in 50 mM Na-MOPS pH 7, 1 mM guanosine and 10 mM Mg^{2+}. 5'radiolabeled oligonucleotide substrate (*$CCUCUA_5$) was added to

initiate cleavage in the presence of 10 mM Mg^{2+}. A time course of substrate cleavage was measured and the data was fit to a double exponential equation where the cleavage rate constant of the burst phase monitors the substrate cleavage by the native ribozyme. The observed rate constant for the first phase was then plotted against each concentration of P5abc, and then fitted to a hyperbolic binding equation: $[S]/ K_d + [S]$.

4.3 RESULTS

4.3.1 Rationale and design of mutant ribozymes.

In the *Tetrahymena* ribozyme, peripheral elements form tertiary contacts P13, P14 and L9/P5 which allow the peripheral element to completely encircle the core helices. Here, we probe whether the physical origin of CYT-18 specificity towards the native state is dependent on a complete ring of peripheral contacts by testing the importance of the long range tertiary contact P14. In this study, two mutations were made to disrupt any potential formation of P14. Both mutants were made in the context of the $E^{\Delta P5abc}$ ribozyme. The L2 single-stranded loop sequence was changed to "UUCG" to disrupt any base pairing interaction with L5c. This mutant will be referred to as the L2 mutant. A second mutant, ΔP2, was made by shortening the P2 stem by 8 base pairs. The rationale behind this mutant was to test if potential interaction by CYT-18 was specific to the RNA backbone.

4.3.2 Catalytic activity of mutants and activation with P5abc or CYT-18

Catalytic activity continues to be an integral method for following the folding of ribozymes and for determining the population of ribozyme that have correctly folded to

the native state. Here, we utilize this method to first test whether the mutant ribozymes are sufficiently active such that folding could be monitored. To assess the fraction of native ribozyme formed with CYT-18, the cleavage rate constant should be faster than the dissociation of substrate from misfolded ribozyme (0.02 min^{-1}) and overall folding to the native state.

Table 4.1 gives a summary of the cleavage rate constants for wt $E^{\Delta P5abc}$, L2 and the ΔP2 deletion mutants. Since folding will be monitored by cleavage reactions in the presence of CYT-18, the cleavage rate constant with CYT-18 was first monitored. For L2 mutant, the cleavage rate constant at 10 mM Mg^{2+} was about 0.6 min^{-1}, very similar to that obtained for the wt $E^{\Delta P5abc}$ (0.8 min^{-1}). In both cases, the cleavage rate is much faster than substrate dissociation and folding so that the amplitude of the cleavage burst would indicate the fraction of native ribozyme present. Interestingly, the cleavage rate constant for the ΔP2 mutant in the presence of CYT-18 was 0.02 min^{-1}, much slower than both the wt $E^{\Delta P5abc}$. However cleavage at higher Mg^{2+}, ΔP2 gave a cleavage rate constant of 0.3 min^{-1}. Though this rate constant is 10-fold lower than the L2 mutant under the same conditions, it is still fast enough to be used in activity assays to assess the fraction of native ribozyme during folding.

Other interesting results were observed from the initial characterization of the L2 and ΔP2 mutants. In addition to measuring the cleavage rate constants in the presence of CYT-18, the cleavage rates of the ribozymes alone and the catalytic activation upon addition of P5abc in *trans* at two different Mg^{2+} concentrations were also monitored. At 10 mM Mg^{2+}, the cleavage rate constant of both the L2 mutant and the wt $E^{\Delta P5abc}$ alone were similar, while the cleavage rate constant for ΔP2 was impaired by about 1000-fold

relative to the wt $E^{\Delta P5abc}$. When a saturating concentration of P5abc was added in *trans*, both mutants gave a 1000-fold activation relative to the ribozyme alone (mutant ribozymes required a much higher concentration of P5abc than for wt $E^{\Delta P5abc}$ as described in section 4.3.3 and 4.3.6 below). However, though the overall increase in activity with P5abc was the same for both mutants, the cleavage rate constant for the ΔP2 mutant was still 100-fold slower than the L2 mutant. In addition, with wt $E^{\Delta P5abc}$, P5abc was better able to promote catalysis than CYT-18, giving a 15-fold difference in cleavage rate constants. Interestingly, this was not observed for both mutants as the cleavage rate constant with P5abc were the same within error as with CYT-18. Therefore in the absence of P14, CYT-18 is just as effective as P5abc and highlight futher the importance of the P14 long range contacts for P5abc.

4.3.3 Activity assays to monitor folding of mutant ribozymes

In order to test if CYT-18 is able to directly stabilize the native state, folding of the mutant ribozymes was first monitored in the absence of CYT-18 to establish an equilibrium point. Folding was initiated at 10 mM Mg^{2+} and monitored at varying times from then. At various points, further folding was blocked by the addition of 50 mM Mg^{2+} and P5abc [15], which also activates the ribozyme for catalysis.

The initial fraction of native ribozyme that folded directly to the native state for the L2 mutant was slightly larger than for wt $E^{\Delta P5abc}$. This presumably reflected a slightly larger partitioning fraction between the native and misfolded state, suggesting that this mutant had an altered folding pathway, where a slightly larger fraction avoids misfolding. The L2 ribozyme approached an equilibrium endpoint of 0.56 ± 0.02 with a rate of 0.06

min^{-1}, with the remaining fraction presumably reflecting the presence of a long-lived misfolded state (Figure 4.2A). We also followed the approach to equilibrium for the ΔP2 mutant and found that this mutant folded to an endpoint of 0.4 ± 0.02 with a refolding rate of 0.03 min^{-1}, essentially identical to the wt wt $E^{\Delta P5abc}$ despite the strong deleterious effect of this mutation on catalysis (Figure 4.2B).

4.3.4 CYT-18 stabilization of the native state by ribozyme activity assays

The results of the activity assays above indicated that the L2 and ΔP2 mutant ribozymes equilibrated with endpoints reflecting 56% and 40% native ribozyme respectively. If CYT-18 does preferentially bind to the native ribozyme, then the stabilization conferred would produce an overall increase in the population of native ribozyme relative to the misfolded state as previously demonstrated for wt $E^{\Delta P5abc}$ (chapter 3). Ribozymes were incubated to generate equilibrium mixtures of native and misfolded ribozymes after which CYT-18 was added and the approach to a new equilibrium was monitored. For the L2 mutant, the fraction of native ribozyme increased, implying that CYT-18 bound tighter to the native state. The observed endpoint in the presence of CYT-18 was similar to that previously observed for $E^{\Delta P5abc}$. For ΔP2, however, the equilibrium endpoint with CYT-18 was much lower than that observed for wt $E^{\Delta P5abc}$ (0.61 ± 0.02 vs 0.8 ± 0.02) (Figures 4.3 A, B), initially suggesting that the extent to which CYT-18 stabilzes the native state relative to the misfolded state is not large.

Since the L2 mutant generated native ribozyme to levels comparable to the wt $E^{\Delta P5abc}$, a control experiment was performed to confirm that CYT-18 was indeed

responsible for shifting the equilibrium to favor the native state. A large fraction of native L2 ribozyme was generated by an extended incubation with CYT-18. The fraction of native ribozyme generated was similar to the endpoint obtained when folding was monitored in the presence of CYT-18 (~ 0.8). Upon proteinase K treatment, the fraction of native ribozyme decreased to an endpoint that reflected the same equilibrium point achieved in the absence of CYT-18, thus confirming that CYT-18 is responsible for the stabilizing the native state (Figure 4.4, D.Kung, unpublished).

Both the L2 and ΔP2 mutant were also incubated with P5abc to assess the maximal amount of native ribozyme generated. Interestingly, the concentration of P5abc used previously for wt $E^{\Delta P5abc}$ (500 nM), did not produce an increase in the fraction of native ribozyme beyond the equilibrium value, suggesting that P5abc was not bound in the ground state for these mutants. The concentration of P5abc was then increased to achieve a saturation point. For the L2 mutant, P5abc concentrations greater than 5 μM gave a maximal fraction close to what was observed with CYT-18 and also similar to the fraction obtained by incubating P5abc with wt $E^{\Delta\ P5abc}$ (Figure 4.5 A). This result implies that the CYT-18 was able to shift the equilibrium towards native to a point where misfolded ribozyme was undetectable (dotted line, Figure 4.3).

Interestingly, the fraction of native ribozyme achieved with ΔP2 mutant in the presence of P5abc was similar to that obtained with CYT-18. Both these endpoints however, were much lower than that previously observed for the wt $E^{\Delta P5abc}$. The low endpoints would initially suggest that the remaining fraction of ribozyme population is damaged. However, experiments performed with independent preparations of P5abc and ΔP2 RNAs ruled out this possibility. Another possibility could be that a significant

fraction of the ribozyme is trapped in some conformation that is not perturbed by the presence of either P5abc or CYT-18. In this case, the endpoint achieved with CYT-18 appears to indicate a complete shift in the equilibrium point to favor the native state.

4.3.5 Binding and dissociation of CYT-18 to ribozyme mutants

For a more quantitative approach to CYT-18 stabilization, the kinetic parameters of binding of CYT-18 to the native and misfolded ribozymes from both mutants were measured. In addition to the ΔP2 and L2 mutants, loop L9.1 in wt $E^{\Delta P5abc}$ was mutated to "UUCG" to disrupt the long range contact P13. This mutant will be referred to as L9.1 (Figure 4.1). P5abc was previously shown to favor the formation to the native state of the L9.1 mutant 100-fold more than the misfolded ribozyme giving 4.1 kcal/mol of stability to the native state (T.Johnson & R.Russell unpublished).

A pulse-chase procedure, as explained in chapter 3, was used to measure the dissociation of CYT-18 from the native and misfolded ribozyme mutant ribozymes. Results showed that CYT-18 dissociates with a rate constant of $(5.6 \pm 0.4)x10^{-2}$ min^{-1} for the L2 mutant, $(3.6 \pm 0.1)x10^{-2}$ min^{-1} for ΔP2 mutant (Figure 4.6 A,B), both giving less than 2-fold difference than that measured for wt $E^{\Delta P5abc}$. The L9.1 mutant also displayed about a 5-fold increase in k_{off} (Figure 4.6 C) when compared to wt $E^{\Delta P5abc}$, suggesting that even with peripheral disruption on the other side of the core, CYT-18 is able to form a relatively stable complex with the native ribozyme. Experiments following the dissociation from the misfolded ribozyme showed that, like wt $E^{\Delta P5abc}$, a rapid phase of dissociation was observed. This rapid phase was followed by a slower phase reflecting

dissociation from the small population of native ribozyme formed during the incubation which generated mostly misfolded ribozyme.

Binding rate constants were also measured. CYT-18 binds rapidly with an association rate of (1.2 ± 0.35) x 10^8 M^{-1} min^{-1} for the L2 and 1 x 10^8 M^{-1} min^{-1} for ΔP2 (data not shown). Therefore, the calculated K_d's for CYT-18 binding to the native mutants are 0.4 nM for the L2 mutant and 0.36 nM for the ΔP2 mutant, similar to that previously calculated for wt $E^{\Delta P5abc}$ (0.2 nM). The results presented here clearly show that CYT-18 binds with almost equal affinity to the native ribozyme of the L2 and ΔP2 mutants and furthermore dissociation kinetics suggests that the differences in affinity between the native and misfolded are also significantly large.

4.3.6 Affinity of P5abc to the native mutant ribozymes

Earlier results in section 4.3.4, showed that P5abc was unable to increase the fraction of native ribozyme under the same concentration as used for wt $E^{\Delta P5abc}$ (500 nM P5abc), which suggested that P5abc was not bound in the ground state. In addition, reactions which monitored the total fraction of native ribozyme by activity assays were only achieved with concentration of P5abc ≥ 5 μM, initially suggesting decreased binding affinity to the native state.

A measurement of binding affinity was made through catalytic activity assays. Both mutants were incubated with varied P5abc concentrations (300 nM-10 μM) for sufficient time to ensure binding, after which a time course of substrate cleavage was monitored at 10 mM Mg^{2+}. The cleavage rate constant for the L2 mutant increased to 1.5 min^{-1} while the maximal cleavage rate for the ΔP2 mutant was 100-fold lower. Both data

sets were fit to hyperbolas, giving a measured K_d for P5abc binding to the native state as 1.8 ± 0.2 μM for the L2 mutant and 1.1 ± 0.3 μM for ΔP2 mutant (Figure 4.7 A,B) . A K_d measured at 1.1 μM and 1.8 μM respectively gives at least 10^8-fold weaker binding than the native of $E^{\Delta P5abc}$ (K_d Native $E^{\Delta P5abc}$ •P5abc = 0.06 pM) [58]. These results greatly contrast the effect administered by CYT-18 protein which gave similar binding affinities to the native ribozyme, and support a model whereby the origin of specificity is distinct for P5abc and CYT-18.

4.4 DISCUSSION

Peripheral structure, in particular P5abc, has long been known to play a critical role in stabilizing the native state through the formation of tertiary contacts [56, 58, 142]. The peripheral element P5abc, not only occupies the same binding interface as CYT-18 [63], but is involved in three out of the five long range tertiary contacts that encircle the core. In this chapter, the differences between the strategies used by proteins and peripheral elements in stabilizing the native ribozyme were investigated. The fact that CYT-18 and P5abc both stabilize the native state of wt $E^{\Delta P5abc}$ and occupy the same binding interface provided an opportunity to directly compare their strategies. For P5abc, the formation of a network of peripheral contacts is predicted to be critical. However, it is currently not known if the same physical mechanism also applies to CYT-18.

Mutants L2 and ΔP2 were designed to disrupt formation of a long range peripheral contact made between L5c and L2, termed P14, to test its role in native stability. Both L2 and ΔP2 mutant ribozymes were sufficiently active, and cleavage of the substrate was fast enough to be used in activity assays to monitor folding of the

ribozymes. CYT-18 was able to shift the equilibrium to favor the formation of the native ribozyme for the L2 mutant. The extent to which the ribozyme populated the native state was found to be similar to wt $E^{\Delta P5abc}$ as the fraction obtained with P5abc indicated that essentially all the misfolded ribozyme refolded to the native state. Interestingly, when CYT-18 was added to the ΔP2 mutant, a small shift in the equilibrium was observed to a level much lower than wt $E^{\Delta P5abc}$. When P5abc was added, the maximal fraction of native ribozyme achieved was similar to that observed with CYT-18. Given that the rest of the ribozyme population appears not to be influenced by the presence of either CYT-18 or P5abc, most simply suggests that the extent to which CYT-18 specifies the native state relative to the misfolded state is also large. Furthermore, dissociation kinetics revealed that the ΔP2 mutant, as well as the L2 mutant, displayed large differences between the native and the misfolded state similar to that observed in wt $E^{\Delta P5abc}$, suggesting that the specificity towards the native state relative to the misfolded state for both mutants is maintained to an extent that is similar to the wt $E^{\Delta P5abc}$.

P5abc binding to both mutants was drastically weakened with K_d values of 1.1 μM for ΔP2 and 1.8 μM for L2 mutant for the native ribozymes. Remarkably, these values represent a 10^8 fold decrease in affinity. Not surprisingly, the P14 disruption, which is a direct contact with P5abc, is much more deleterious than the P13 and L9/P5 mutants where the K_d for binding the native state was 0.1 nM and 40 pM respectively (T.Johnson). These results highlight how P5abc and CYT-18 differ mechanistically despite being functionally analogous. P5abc uses a set of long range tertiary contacts formed with other peripheral domains and is critically dependendent on cooperative formation of these contacts to preferentially stabilize the native state, while CYT-18 has

evolved specific insertions, H0, Ins1 and Ins2 which make more local contacts along the P4-P6 domain [72, 77].

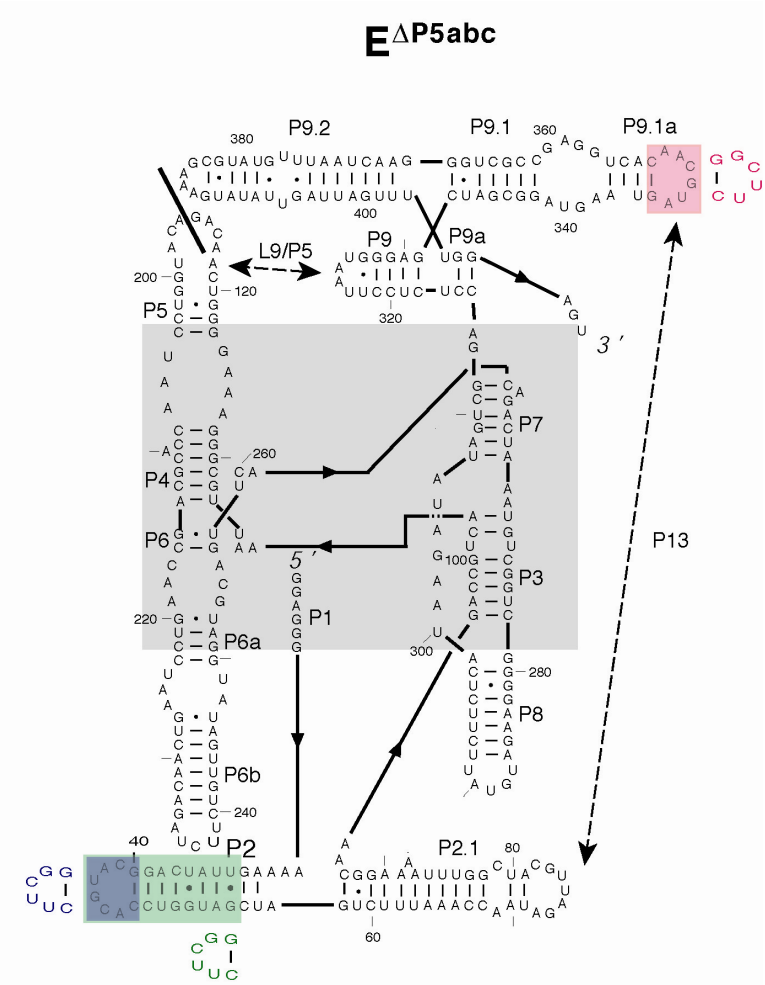

Figure 4.1: Secondary structure map of $E^{\Delta P5abc}$ highlighting changes made by mutagenesis

Regions of mutations are shaded in boxes. Loop 2 sequence was mutated to "UUCG" (blue) to give the L2 mutant and P2 stem was shortened by 8 base pairs and ends in the same "UUCG" loop sequence (green) to give the ΔP2 mutant. Disruption of the base pairing interaction of the P13 long range tertiary contact by mutating loop L.9.1 to "UUCG" are highlighted in pink boxes.

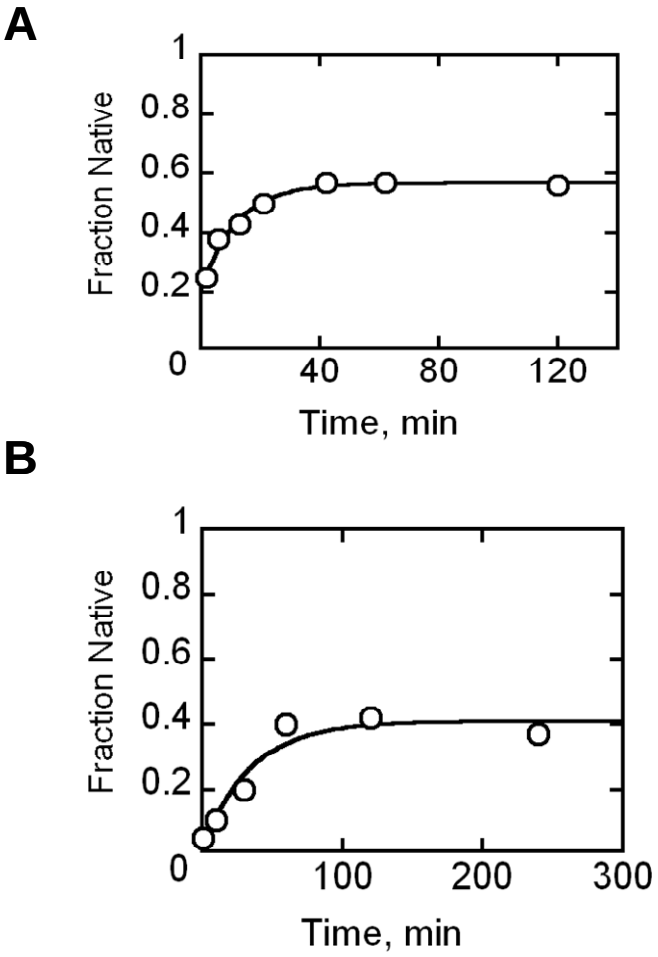

Figure 4.2: Approach to an equilibrium value for ribozyme mutants using a ribozyme cleavage assay

A, L2 mutant refolds at a rate of 0.06 min^{-1} to an equilibrium mixture of 56 % native ribozyme with the rest being either misfolded or damaged ribozyme. The starting point for this mutant lies around 0.2 indicating a small population of ribozyme rapidly folds to the native state. **B**, Refolding of ΔP2 mutant with an observed rate of 0.03 min^{-1} and an end point of 0.4 at equilibrium

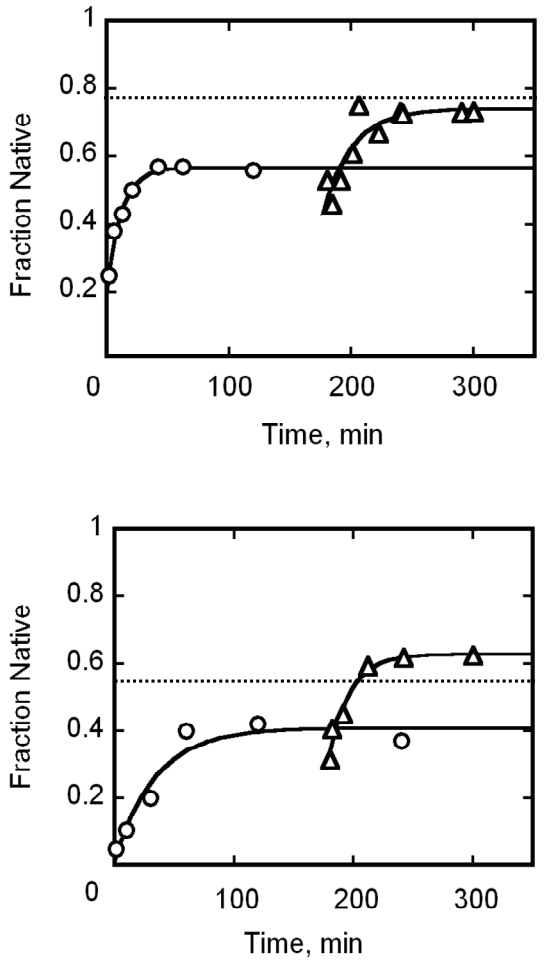

Figure 4.3: CYT-18 shifts the equilibrium to favor the formation of native ribozyme

Addition of CYT-18 to an equilibrium mixture of each mutant ribozyme. The change in the fraction of native ribozyme was monitored at varying times after the addition of CYT-18. Circles follow the L2 and P2 mutants folding to an equilibrium mixture. Arrows indicate the time CYT-18 was added. The time course following the addition of CYT-18 is followed by open triangles. **A**, CYT-18 increases the fraction of native ribozyme for L2 mutant giving approximately 0.73 ± 0.01 native ribozyme. **B**, CYT-18 increases the fraction of the native ribozyme to 0.61 ± 0.02 for ΔP2 mutant. Dotted lines show the fraction of native ribozyme achieved with saturating concentrations of P5abc.

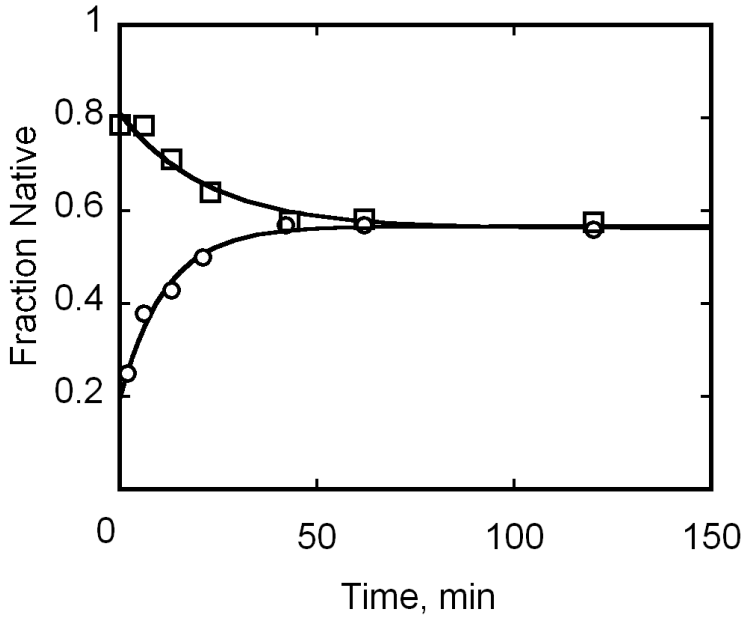

Figure 4.4: Approach to the steady state equilibrium for L2 mutant

The L2 mutant ribozyme equilibrates with ~ 56% native ribozyme. CYT-18 was incubated with the ribozyme to generate essentially 100 % native ribozyme. CYT-18 was then removed by adding proteinase K (1 mg/ml) and the decrease in the fraction of native ribozyme was monitored over time (squares). The fraction of native ribozyme decreased from a fraction on 0.81 to 0.56, the same equilibrium point as achieved in the absence of CYT-18 (circles).

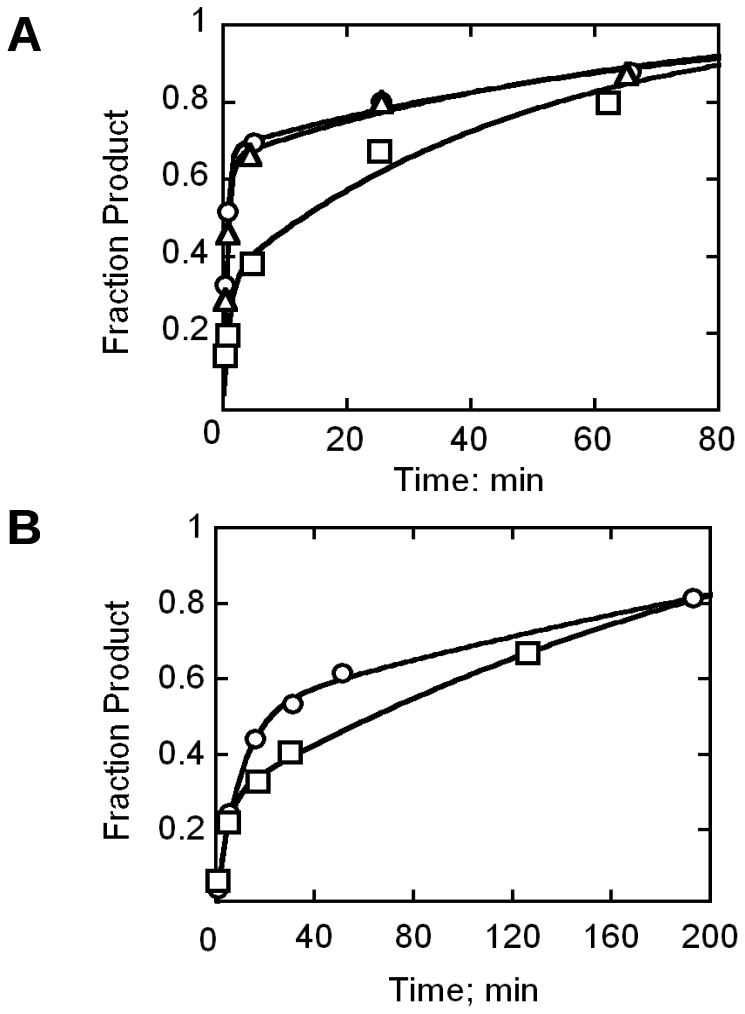

Figure 4.5: Fraction of native ribozyme achieved with saturating concentrations of P5abc

A, Cleavage assays with 200 nM L2 mutant ribozyme in complex with 1 μM P5abc (squares), 5 μM P5abc (circles) and 7 μM P5abc (triangles) at 25 °C, before the addition of radiolabeled substrate to assess the fraction of native ribozyme. Cleavage reactions were performed at 50 mM Mg^{2+}. The fraction of native ribozyme in the burst phase gave 0.72 ± 0.04 at 5 μM and 7 μM P5abc **B**, Cleavage reaction for ΔP2 mutant performed with 1 μM (squares) and 9 μM (circles). The fraction of native ribozyme attained was 0.55 ± 0.04 with 9 μM P5abc.

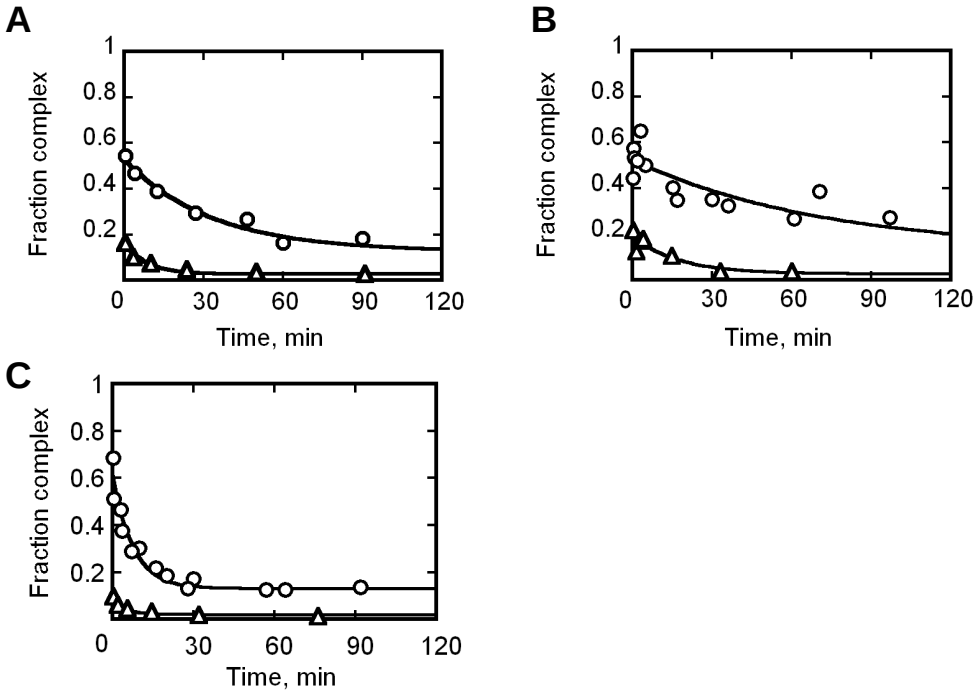

Figure 4.6: Dissociation of CYT-18 from the native and misfolded mutants

L2, ΔP2 and P9.1 mutants gave similar dissociation kinetics for native and misfolded ribozymes. Data was fit to a single exponential equation $(A)*(1-e^{-kt}) + c$, where A is amplitude, k is the observed rate constant for dissociation and c is the endpoint. Dissociation of 200 nM CYT-18 from native ribozymes is followed by circles and dissociation from a small population of native ribozyme formed when measurements are made mainly misfolded ribozyme is followed by triangles. The native complex was formed with CYT-18 (37 °C for 30 min) and dissociation was measured at 25°C upon the addition of 10-fold excess unlabeled RNA (pre-folded 50 °C for 30 min). **A**, CYT-18 dissociates from the native ribozyme of the L2 mutant with a rate constant of (5.6 ± 0.4) x 10^{-2} min^{-1}. Dissociation from the misfolded ribozyme is rapid and complete before 15 seconds with a large fraction of the complex dissociating on this time scale. Following this rapid dissociation the remaining 10-20% of the complex dissociated with a rate constant of 0.1 min^{-1}, possibly reflecting dissociation small population of native **B**, Dissociation of CYT-18 from ΔP2 native ribozyme with a rate constant of (3.6 ± 0.1) x 10^{-2} min^{-1}. CYT-18 also come off rapidy from the misfolded ribozyme with the small fraction of native ribozyme dissociating at 0.05 min^{-1} **C**, Dissociation of CYT-18 from the native L9.1 measured as 0.13 min^{-1}, ~5-fold faster than wt $E^{\Delta P5abc}$ followed by rapid dissocation from misfolded followed by the smaller phase at 0.2 min^{-1}.

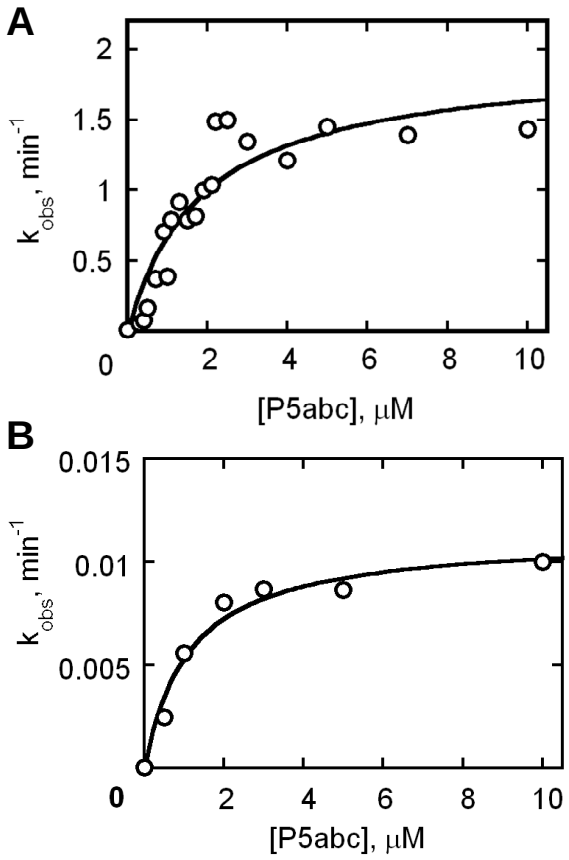

Figure 4.7: Binding of P5abc to mutant ribozymes

A, 200nM L2 mutant was incubated with P5abc (300 nM to 10 μM) at 25° C for 3 hrs. After the addition of *S, a time course of cleavage was observed at 10 mM Mg^{2+} at each concentration of P5abc. Data was fit to a hyperbolic binding equation $k[S_o]/ K_d + [S_o]$ where $[S_o]$ is the free P5abc concentration and Kd is the equilibrium dissociation constant. Fitting the data gave K_d's of 1.8 μM for the L2 mutant (D.Kung, unpublished) **B,** Observed K_d for P5abc binding to ΔP2 was measured at 1.2 μM.

RNA	Substrate cleavage at 10 mM Mg^{2+} (min^{-1})	Substrate cleavage at 50 mM Mg^{2+} (min^{-1})
$E^{\Delta P5abc}$	*0.01	*0.57
$E^{\Delta P5abc}$ • P5abc	7.0 ±1.3	n/d
$E^{\Delta P5abc}$ • CYT-18	0.8 ± 0.2	3.0 ± 0.5
L2UUCG	*0.01	*0.51
L2UUCG • P5abc	1.1 ± 0.2	2.2 ± 0.4
L2UUCG • CYT-18	0.6 ± 0.07	*3.0
ΔP2	1.0×10^{-5}	1.0×10^{-4}
ΔP2 • P5abc	0.01 ± 0.004	0.1 ± 0.015
ΔP2 • CYT-18	0.02 ± 0.001	0.3 ± 0.03

*-value obtained from a single experiment

n/d -value not determined

Table 4.1: Cleavage rate constants for the L2 and P2 mutant ribozymes

Observed rate constants are reported from at least two independent experiments unless otherwise indicated. For reactions with $E^{\Delta P5abc}$, 500 nM P5abc and 500 nM CYT-18 were used. For reactions with L2 mutant and ΔP2 mutant, 1 μM CYT-18 and 5-7 μM P5abc were used which were previously shown to be saturating. Reactions were performed in 10 mM Mg^{2+}, 50 mM Na-MOPS, pH 7. Cleavage reactions were initiated at 25 °C with 1 mM guanosine or 2 mM G for ribozymes alone.

APPENDIX

Mapping of the functional boundaries and secondary structure of the mouse mammary tumor virus Rem-responsive element

Jennifer A. Mertz[‡], Amanda B Chadee[¶], Hyewon Byun[‡], Rick Russell[¶], and Jaquelin P. Dudley[‡]

From the [‡]Section of Molecular Genetics and Microbiology, [¶]Department of Chemistry and Biochemistry, and Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX 78712

Running Title: Mapping of the MMTV RmRE

ABSTRACT

Mouse mammary tumor virus (MMTV) is a complex retrovirus that encodes at least three regulatory and accessory proteins, including Rem. Rem is required for nuclear export of unspliced viral RNA and efficient expression of viral proteins. Our previous data indicated that sequences at the envelope-3' long terminal repeat junction are required for proper export of viral RNA. To further map the Rem-responsive element (RmRE), reporter vectors containing various portions of the viral envelope gene and the 3' long terminal repeat were tested in the presence and absence of Rem in transient transfection assays. A 476 bp fragment that spans the envelope-long terminal repeat junction had activity equivalent to the entire 3'-end of the mouse mammary tumor virus genome, but further deletions at the 5'- or 3'-ends reduced Rem responsiveness. RNase structure mapping of the full-length RmRE and a 3'-truncation suggested multiple domains with local base-pairing and intervening single-stranded segments. A secondary structure model constrained by these data is reminiscent of the RNA response elements of other retroviruses, with numerous local stem-loops and long-range base pairs near the 5'- and 3'-boundaries, and differs substantially from an earlier model generated without experimental constraints. Co-variation analysis provides limited support for basic features of our model. Reporter assays in human and mouse cell lines revealed similar boundaries, suggesting that the RmRE does not require cell type-specific proteins to form a functional structure.

A.1 INTRODUCTION

Mouse mammary tumor virus (MMTV) has multiple regulatory and accessory genes (1, 2). The known accessory genes specify a dUTPase (3), which is believed to be involved in retroviral replication in non-dividing cells (4), as well as superantigen (Sag). Sag is a transmembrane glycoprotein that is involved in the lymphocyte-mediated transmission of MMTV from maternal milk in the gut to susceptible epithelial cells in the mammary gland (5 ,6). The Sag protein expressed by endogenous (germline) MMTV proviruses has been reported to provide susceptibility to infection by exogenous MMTVs or the bacterial pathogen, *Vibrio cholerae* (7). These results suggest a role for MMTV Sag in the host innate immune response.

MMTV recently was shown to be a complex retrovirus (1). Complex retroviruses encode RNA-binding proteins that facilitate nuclear export of unspliced viral RNA by using a leucine-rich nuclear export sequence (NES) (8), which binds to chromosome region maintenance 1 (Crm1) (9), whereas simple retroviruses have a cis-acting constitutive transport element (CTE) that directly interacts with components of the Tap/NXF1 pathway (10). Similar to other complex retroviruses, MMTV encodes a Rev-like protein, regulator of export/expression of MMTV mRNA (Rem) (1). Rem is translated from a doubly spliced mRNA into a 33 kDa protein that contains nuclear and nucleolar localization signals as well as a predicted RNA-binding motif and leucine-rich NES (1 ,2). Our previous experiments indicated that Rem affects export of unspliced viral RNA, and a reporter vector that relies on luciferase expression from unspliced RNAs has increased activity in the presence of Rem (1). Sequences at the MMTV envelope-long terminal repeat (LTR) junction were required within the vector for Rem-induced expression, suggesting that the LTR contains all or part of the Rem-responsive element

(RmRE). Very recently, Mullner et al. identified a 490-nt region spanning the MMTV envelope-3' LTR region, which was predicted to form a highly structured RNA element (11). This element confers Rem-responsiveness on heterologous human immunodeficiency virus type 1 (HIV-1)-based plasmid constructs in transfection experiments.

Experiments using other retroviral export proteins have demonstrated considerable variation in the size of the response elements. A minimal Rev-responsive element (RRE) in the human immunodeficiency virus type 1 (HIV-1) genomic RNA is 234 nt, the human T-cell leukemia virus (HTLV) Rex-responsive element (RxRE) is 205 nt (12-14), whereas the Rec-responsive element (RcRE; also known as the K-RRE) of human endogenous retrovirus type K (HERV-K) is 416 to 429 nt (15 ,16). Most response elements are confined to the 3' end of their respective retroviral genomes (either to the envelope or LTR regions) (14 ,15), but 5' Rev-response elements also have been identified (17). Studies indicate that secondary structure is a critical factor for proper function of retroviral response elements (18), and that multiple stem-loops are required. Export proteins multimerize on these elements to allow activity (19).

In the current study, we have used deletion mutations within a reporter vector based on the 3' end of the MMTV genome to define a 476 nt element necessary for maximum Rem responsiveness. This element spans the envelope-LTR junction of the MMTV genome as previously reported (1). However, a secondary structure model generated using digestions of the RmRE by RNases V1, T1 and A as experimental constraints differs significantly from the published structure (11) and more closely resembles complex retroviral response elements. Transfection experiments indicated that the MMTV RmRE could function in both mouse and human cells, suggesting that conserved cellular proteins interact with Rem.

A.2 EXPERIMENTAL PROCEDURES

A.2.1 Cell cultures and tranformation

XC rat fibroblast cells were cultured in Dulbecco's modified Eagle's medium supplemented with 5% fetal calf serum (FCS), gentamicin sulfate (50 μg/ml), penicillin (100 U/ml) and streptomycin (50 μg/ml). Jurkat human T lymphoma cells were maintained in RPMI media supplemented with 5% FCS, gentamicin sulfate (50 μg/ml), penicillin (100 U/ml) and streptomycin (50 μg/ml). XC cells were transfected in 6-well plates using DMRIE-C transfection reagent (Invitrogen) according to the manufacturer's instructions. Each transfection contained 250 ng of the reporter vector, 250 ng of pGL3-control plasmid, and 2 μg of an expression vector for Rem or green fluorescent protein (GFP). Jurkat cells were transfected by electroporation using a BTX ECM600 instrument. Cells (1 X 10^7) were mixed with the appropriate plasmid DNA in a volume of 400 μl Jurkat medium prior to electroporation in 4 mm gap cuvettes (260 V, 1050 μF and 720 ohms). Each transfection contained 1 μg of the reporter vector, 1 μg of the GL3-control plasmid, and 20 μg of expression vectors for Rem or GFP. Transfected cells were incubated at 37 °C in complete medium and harvested two days after transfection. All transfections were performed in triplicate using the same amount of total DNA in each sample. Cytoplasmic extracts obtained by three freeze-thaw cycles were stored at -70 °C prior to luciferase assays.

A.2.2 Plasmid constructs

The plasmids GFP-Rem (RemP71), HM*Rluc* and HMΔ*e*LTR*luc* have been described elsewhere (1). The plasmid pEGFPN3 was obtained from Clontech and the pGL3-control plasmid was obtained from Promega. The mutant HMΔ*e*LTR+X*Rluc*

constructs were made by insertion of the mutant RmRE into an engineered *Sca*I site downstream of the splice acceptor site and upstream of the SV40 poly (A) signal in HMΔ*e*LTR*luc* (Fig.A1). *In vitro* transcription vectors for wild-type and mutant RmREs were made by PCR amplification of the wild-type or mutant RmRE with insertion of a T7 polymerase promoter upstream of the RmRE. The PCR product was then inserted into the multiple cloning site of the pEGFPN3 vector. The 1-348 mutant containing the 3' truncated RmRE was amplified with the following primers: T7HM+ 5' TAA TAC GAC TCA CTA TAG GGA TCT TAA CGT GCT TC 3' and RmRE 348- 5' AGT ACT GTG GTC CTT GCC TCA GGA GG 3'. Cloning of the GFP-RemP71L expression vector and all other mutant RmRE constructs was performed using site-directed mutagenesis or by cleavage with restriction enzymes and religation. Details are available upon request. All constructs were confirmed by automated sequencing reactions.

A.2.3 Luciferace assays

Luciferase assays were performed using the Dual Luciferase Reporter Assay System (Promega) to quantitate both Renilla and firefly luciferase activities (20). A firefly luciferase reporter vector lacking MMTV sequences was added to each transfection, and activities showed that different transfections within the same experiment had similar DNA uptake. Samples of normalized and unnormalized data have been provided (see Table S1 compared to Fig.A2).

A.2.4 Western Blotting

Protein extracts for Western blotting were obtained as previously described (1). Western blots were performed using antibodies specific for the GFP tag (Clontech) on GFP-Rem or actin (Calbiochem); the latter served as a control for protein loading. Proteins were detected using the ECL Western blotting detection system (Amersham).

A.2.5 In vitro transcription and end-labelling of RNAs

DNAs encoding the wild-type and truncated RmREs (1-496 and 1-348) were linearized using ScaI (New England Biolabs). RNAs were prepared by in vitro transcription using T7 RNA polymerase and purified using a Qiagen RNeasy column following the manufacturer's protocol. RNAs were incubated with shrimp alkaline phosphatase (Promega) to remove the 5'-triphosphate and then with [γ-^{32}P] ATP (Perkin Elmer) and T4 polynucleotide kinase (New England Biolabs) to add a radiolabeled phosphoryl group to the 5'-end of the RNA (21). RNAs were 3'-end-labeled using [5'-^{32}P] pCp and T4 RNA ligase to extend the 3' end of the RNA with a single labeled nucleotide (22). Both 3'- and 5'-labeled RNAs were purified using 8% native polyacrylamide gel electrophoresis (PAGE).

A.2.6 Nuclease Mapping of RmRE RNAs

RNase footprinting reactions (10 µl) were performed using in vitro-transcribed full-length (1-496) or 3'-truncated (1-348) RmRE RNA. Labeled RNA (3 – 15 nM, 10,000 cpm/µl) was incubated in 50 mM sodium MOPS (pH 7.0) and 10 mM MgCl2. Varying the pre-incubation time did not affect the results (15 sec – 5 min), suggesting that the RNA formed secondary structure rapidly and stably. Nucleases were then added at levels empirically determined to give optimal RNA cleavage: 0.1 U of RNase T1 (Sigma), 10^{-4} U of RNase V1 (Ambion) and 10^{-3} U of RNase A (Sigma). RNA was digested for 1 to 3 min at 25°C, and aliquots were quenched by adding 1 mg/ml proteinase K and then two volumes of 20 mM EDTA in loading dye [90% (v/v) formamide, 0.04% xylene cyanol and 0.04% bromophenol blue]. Reaction tubes were immediately placed in liquid nitrogen to ensure that the RNases were fully inactivated. RNA sequencing ladders were generated by digesting 5'- or 3'-labeled RNA with 0.1 U of RNase T1 for 15 min under denaturing conditions (7 M urea at 50°C). Reaction

products were separated by 8% denaturing PAGE and quantitated using Semi-Automated Footprinting Analysis software (SAFA) (23). From each digestion, raw intensity values corresponding to cleavage at each nucleotide were normalized by dividing each value by the average intensity of all bands within the range quantitated. These normalized values were then averaged between experimental determinations to produce final values. Except where indicated, all reported results reflect the averages of two to six independent determinations.

A.3 RESULTS

A.3.1 RmRE maps to a large region spanning the junction of the MMTV envelope gene and the 3' LTR

To determine the boundaries of the MMTV RmRE, several constructs were designed based on our previously described pHMRluc vector (1, 24). This vector contains the 3' end of the MMTV genome, including part of the envelope gene and the 3' LTR, downstream of the cytomegalovirus (CMV) promoter (Fig.A1). Since the Renilla luciferase gene was inserted between the splice donor and acceptor sites in the envelope gene, detection of luciferase activity in transfected cells indicates export of unspliced mRNA from the nucleus to the cytoplasm. Previous data showed that reporter gene activity from this vector is induced by co-expression of the MMTV export protein, Rem, in trans (1). Rem-induced luciferase activity also required the presence of the envelope-3' LTR junction in the reporter vector, suggesting that these sequences contain the RmRE (1). The pHMΔeLTRluc plasmid, which substitutes the simian virus 40 (SV40) polyadenylation region for the MMTV 3' LTR (Fig.A1), shows no response to the addition of Rem expression vectors (1). Therefore, various MMTV sequences at the

envelope-LTR border were re-inserted into the pHMΔeLTRluc vector to determine the region necessary for Rem responsiveness.

Our previous experiments showed that insertion of a BglII to ScaI fragment (496 bp) within the pHMΔeLTRluc plasmid gave the same Rem response in mouse cells as the wild-type pHMRluc vector, which contains the entire 3' end of the MMTV genome (1). In agreement with these results, both reporter plasmids showed ca. 5-fold increases in luciferase activity after co-expression of Rem in transient transfections of XC rat cells (compare pHMRluc to the 1-496 vector in Fig. A2-A). Unlike mouse cells, XC rat cells lack endogenous Mtv proviruses, which may express virally encoded proteins (7). Western blotting also was used to verify expression of the Rem construct (data not shown).

Subsequent plasmids were designed to determine the 5' border of the response element (RmRE). Deletion of 10 or 20 nt from the RmRE 5' end had no reproducible effect on Rem responsiveness (constructs 11-496 and 21-496) (Fig.A2-A). However, reporter plasmids with a deletion of 30 bp from the 5' end of the vector showed a 2.5-fold decrease in reporter levels after Rem expression, and constructs with a deletion of 40 bp or more had little or no detectable Rem response (Fig.A2-A).

Deletions also were performed to remove sequences from the 3' end of the RmRE. Removal of 98 or 123 bp at the 3' end of the RmRE of the reporter vector resulted in a 2-fold drop in Rem responsiveness, and deletion of 148 bp essentially showed no Rem response (Fig.A2-B). Furthermore, a reporter construct carrying an internal deletion of the RmRE between 50 and 369 bp (Δ50-369) had little detectable Rem response. These results suggest that full Rem responsiveness requires the majority of the BglII-ScaI fragment spanning the envelope-3'LTR region.

A.3.2 MMTV RmRE has extensive and complex secondary structure

After defining the functional boundaries of the RmRE, we used RNase mapping to probe its secondary structure. RNA spanning the wild-type C3H MMTV RmRE (nt 1-496) was produced by in vitro transcription, 5'- or 3'-end-labeled with ^{32}P, and subjected to limited digestion with RNases T1, A, and V1. RNase T1 cleaves preferentially at single-stranded G residues, RNase A cleaves at single-stranded C or U residues, and RNase V1 primarily digests base-paired nucleotides without strong preferences for nucleotide identity (25). Thus, RNase probes used in combination give extensive information on base-pairing status throughout a structured RNA (26, 27).

Results from digestions with each RNase are shown for a portion of the RmRE (Fig.A3). The RNA includes extensive secondary structure, as indicated by localized regions that were cleaved efficiently by RNase V1 and were inaccessible to RNases A and T1 (e.g., nt 199-200 and 231-234) (Fig.A3-A). On the other hand, these segments were limited to a few consecutive nucleotides and were interrupted by local regions of accessibility to the single-strand specific RNases, suggesting a structure in which short helical segments are separated by many hairpin and/or internal loops.

To evaluate the RNase mapping results quantitatively, we used the freely-available software SAFA (23) to determine the intensity of RNase-mediated cleavage at each position. Digestion profiles for the three RNases are provided (Fig.A3-B) across the range shown in the gel (Fig.A3-A). Profiles for the entire RmRE also are shown (Figs. S1-S3). From the average profiles of multiple independent determinations, we established empirically two threshold levels of intensity. Bands that significantly exceeded the average band intensity were considered to reflect nucleotides that were accessible to each nuclease. A second, higher threshold was also established to separate the smaller groups of nucleotides that were the most accessible to each RNase probe (Fig. S1-S3).

With the latter set of the most reproducible and largest signals as experimental constraints, we used Mfold to generate possible secondary structures of the RmRE (28). The structure predicted to be most stable, after using constraints from the RNase digestions, is shown (Fig.A4). This structure is highly complex, containing many single-stranded regions and multiple hairpin loops typical of the response elements of other complex retroviruses (12, 26, 29-31). This structure has been divided into regions I, II, III, and IV. Also similar to other response elements, the secondary structure models favored by Mfold include long-range base-pairings between sequences close to the 5'-and 3'-boundaries of region I of the RmRE within region I (Fig.A4 and data not shown).

Interestingly, in the absence of the experimental constraints, Mfold predicted secondary structures that differed substantially from that shown in Fig.A4. The most favorable structure in the absence of the constraints is shown for comparison (Fig.S2). Although this structure retains the most basic features of the model in Fig.A4 –multiple stem-loops and long-range pairings between sequences near the boundaries – local interactions differ substantially. Importantly, the model generated by including experimental constraints provides better agreement with the footprinting data that were not used as constraints. Of 43 nucleotides that were reactive to RNase V1, but were not constrained in the structural prediction, 72% are correctly predicted to be double-stranded, compared with only 35% for the unconstrained model (data not shown). The agreement with RNase A data was comparable for the two models (41% of reactive nucleotides are single-stranded in the model generated with constraints versus 35% for the model without constraints), whereas, for RNase T1, only six reactive nucleotides were not used as constraints, preventing meaningful analysis.

When all of the footprinting data are considered, the constrained model gives much higher rates of agreement for all three RNases. While this is true by necessity

because many of the reactive nucleotides were constrained in the modeling, this comparison nevertheless highlights the point that a secondary structure may be obtained to give much stronger agreement with experimental results than the one that would be favored in the absence of experimental constraints.

A.3.3.The limits of the MMTV RmRE are unaffected by cell type

Previous experiments suggest that Rem is synthesized as a 33 kDa precursor protein, which is targeted by an unusually long signal peptide (SP) (98 amino acids) for translocation across the endoplasmic reticulum (ER) membrane (1, 32). Subsequently, the Rem C-terminus is glycosylated, and the N-terminal SP appears to be cleaved by signal peptidase in the ER lumen. Our previous data showed that GFP-tagged Rem accumulates in nucleoli (1). Experiments by Dultz et al. (32) showed that cleaved Rem SP is released into the cytoplasm prior to nuclear localization. We have observed similar results using cells transfected with Rem cDNA tagged on the N-terminus with GFP; the major product obtained was consistent with the size of GFP plus the Rem SP (~38 kDa) (Fig.A5-A). Furthermore, GFP-tagged Rem constructs with a leucine instead of proline at position 71 (P71L) had greater induction of reporter activity with the 1-496 construct (~24-fold) (Fig.A5-B) compared to RemP71 (5 to 7-fold) in XC rat cells. Virtually identical results were obtained with untagged RemP71L (data not shown). Previous data indicated that RemGFP (P71) is poorly cleaved to SP (1), suggesting that SP is the active form of the protein in the reporter assay.

GFP-tagged RemP71L then was tested with 5’ deletion mutants of the RmRE. As noted with the less active version of the protein (Fig.A2), deletion of 30 bases from the 5’ end of the RmRE was responsive to Rem. Although the basal activity of the 31-496

mutant was approximately 5-fold lower in some assays, the induction by RemP71L was similar to that observed for the wild-type construct (1-496). The 31-496 mutant showed about 50% of the wild-type activity using RemP71, which may be a consequence of proline instead of leucine at position 71. However, deletion of another 10 bases from the 5' end of the RmRE removed all Rem responsiveness (mutant 41-496).

Nucleotides 30 to 40 are essential, either because they form an important structure or because their deletion leads to larger scale rearrangements. Mutants deleted at the 3' end of the RmRE also were tested for their response to RemP71L. The results indicated that the 1-373 mutant had ca. 50% of the Rem responsiveness of the wild-type RmRE, whereas the 1-348 mutant had approximately 25% of wild-type RmRE activity. These experiments also agree with those obtained with RemP71 and suggest that deletions at the 3' end of the RmRE are more easily tolerated than those at the 5' end, perhaps because 3' deletions are less likely to give global rearrangements. Consistent with this interpretation, RNase footprinting of the 1-348 mutant showed substantial rearrangements of the 3'-portion of the RmRE but appeared to preserve an extensive local structure element in the 5'-half of the RmRE (nt 70-160, largely region II in Fig. A4).

To determine whether cellular factors would affect the region required for Rem responsiveness, the same 5' and 3' deletion mutants were also tested in human Jurkat T cells (Fig.A5-C). Like XC fibroblasts, Jurkat cells lack endogenous MMTV proviruses that might contribute to Rem activity, yet these cells are capable of infectious virus production (33). Induction by Rem of the 1-496 construct observed in Jurkat cells was higher than that observed in XC cells, which may be due to the higher transfection efficiency of Jurkat cells. However, both the 5' and 3' deletions had a similar response to

RemP71L in each cell type (compare Figs. A5-B and A5-C). These results suggest that the RmRE structure is intrinsic to the RNA sequence and not dependent on the presence of cell-type specific proteins.

A.3.4 Co-variation analysis reports a complex structure with multiple stem loops.

To obtain further biological evidence for the proposed RmRE structure, we performed co-variation analysis on nucleotide changes observed in various MMTV isolates available in public databases. Although the number of such complete sequences is relatively limited, an alignment of the RmRE regions of milk-borne MMTVs and endogenous proviruses that also can be transmitted through the milk [Mtv2 (aka GR-MMTV) and Mtv1) was obtained (Fig.A6). Interestingly, four of the 16 sequences showed a C to U transition at nucleotide 90 that was accompanied by G to A change at nucleotide 95 (C3H/HeJ, SW, JYG, and BALB2) (bold type), its predicted base-pairing partner. Further, most sequences retained the ability to form a wobble base pair by including U at position 90 and G at position 95. In addition, four sequences with a G to A change at position 136 showed a C to U change at position 166 [CS, TES2, RIII/Sa1 and RIII/Sa2 (two isolates)] (bold and underlined type). Positions 136 and 166 are base paired in the unconstrained model (Fig.S2). Only two sequences showed 136A and 166C (no predicted base pairing). No co-variation was observed in the extended hairpin model proposed for the RmRE by Mullner et al. (11). Further, a 3 nt insertion in the GR-MMTV RmRE mapped to a single-stranded portion of region IV. Thus, co-variation analysis provides limited support for both the constrained and unconstrained complex stem-loop models for the MMTV RmRE (Fig.A4), but not the extended hairpin model.

A.4 DISCUSSION

Previous experiments indicated that MMTV is a complex retrovirus that encodes the Rem regulatory protein from a doubly spliced mRNA (1, 2). Rem is required for optimal export of full-length MMTV RNA from the nucleus (1), and also for a post-export function that depends on sequences within the 3' MMTV LTR (24) as well as the junction with the envelope gene (11, 24). In this study, we have mapped the limits of the MMTV RmRE using a reporter vector based on the 3' end of the MMTV genome (1, 24). Although essentially the entire 496 bp RmRE region defined earlier was required for full activity, deletion of 30 bp at the 5' end or up to 148 bp at the 3' end gave constructs that retained partial activity in the assays.

To understand the nature of the RmRE structure, we subjected *in vitro* transcribed RNA to mapping with three RNase enzymes, which recognize different single-stranded or double-stranded nucleotides. The RNase protection data were then used as constraints in Mfold, and a secondary structure prediction was generated. Interestingly, the predicted structure (Fig.A4) is quite different than that previously published for the MMTV RmRE (11). The published RmRE structure, which was solely based on the Mfold algorithm (11, 28), contains predominant double-stranded regions typical of CTEs found in simple retroviruses (34-37). In contrast, our constrained structure has many single-stranded regions and a complex stem-loop structure that more closely resembles the response elements of the complex retroviruses, HIV, HTLV, and HERV-K (13, 16, 26, 30, 38). Recent work has shown that the HIV Rev and HTLV Rex regulatory proteins can function on the MMTV reporter vector pHM*Rluc* in human cells (24). The activity of the heterologous regulatory proteins on the vector required the presence of the RmRE. Thus,

common features of RNA structure appear to be recognized for Rem, Rev, and Rex function. Both Rev and Rex require a stem with a bulge for initial RNA binding followed by multimerization of the protein on the response element (39, 40).

The secondary structure predicted with the inclusion of the structure mapping data is also significantly different from a prediction for the identical region in the absence of the experimental constraints. For RNAs of this size and complexity, experimental constraints are important aids in predictions of secondary structure. Even with the experimental constraints, the accuracy of the secondary structure prediction is limited. First, not all of the nucleotides are constrained by the data, and many possible structures remain compatible with the experimental results. There are also limitations in structure prediction algorithms, most notably the inability to predict pseudoknots. Further, the structure of the RmRE may be influenced by its flanking sequences, by co-transcriptional folding, or by other features of the cellular environment that are not captured *in vitro* in structure mapping experiments. Nevertheless, the central features of the predicted structure the extensive secondary structure, multiple stem-loops, and long-range base pairs are supported by the boundary mapping experiments and are consistent with results from other complex retroviruses (15, 16, 18, 26), suggesting that these features are likely to be present in the natural context of the RmRE.

Consistent with a structural model involving long-range base pairs between the boundary regions, extensive deletion from either the 5'- or the 3'-ends to eliminate the long-range interactions severely compromised Rem-responsiveness in reporter assays (see Figs. A2 and A5 and constructs 55-496 and 1-348). On the other hand, the effects of 5'- and 3'-deletions were not equivalent. Whereas even small deletions from the 5'-end

had strong effects on Rem responsiveness, more substantial deletions were tolerated from the 3'-end. For example, the 5'-deletion mutant 41-496 blocked function, but a deletion from the 3' end, 1-373, retained significant function, even though both deletions are predicted to disrupt the same set of long-range interactions (Fig.A4). Even the more extensive deletion, mutant 1-348, retained activity in some contexts (Fig.A5). The lower sensitivity to 3'-deletions may reflect, in part, a predicted local element toward the 3'-end, which is apparently not essential for function. Alternatively, deletions from the 5'-end may lead to more extensive rearrangements in the central region, and these regions are critical for Rem responsiveness. Supporting this idea, a construct that retained the capacity to form the long-range interactions, but lacked the central region (an internal deletion of bases 50 to 369), had little Rem-responsiveness in reporter assays.

The structural consequences of the 3' deletion mutant, 1-348, were determined. This mutant, which was minimally functional in the reporter assays, was used for RNase mapping and compared to the full-length 496 nt sequence. Quantitation and comparison of RNase VI protection assays revealed that the mutant and wild-type RmREs differed dramatically from nt 160 through the 3' ends (Fig. S1C). Dramatic changes also were observed between nucleotides 240 to 290 using RNase T1 (Fig. S1A) as well as between nt 50 to 60 and between nt 220 and 280 using RNase A (Fig. S1B). Since the reporter assays indicated that the 1-348 mutant had low levels of Rem responsiveness and that the 5' end of the RmRE must remain intact for function, the RNase protection data confirms that the majority of the 5' end of the 1-348 sequence retains wild-type structure. No tertiary structure has been detected for the RmRE in the absence of Rem protein (A. Chadee, unpublished). However, recent experiments have used SHAPE analysis to map

the secondary structure of the HIV RRE in the presence of Rev compared to a mutant RRE selected in the presence of a Rev NES mutant (30). These results indicated that the RRE has considerable flexibility in structure to allow Rev binding and function, suggesting that the 3' end of the RmRE also may have a flexible structure.

Current reports indicate that Rev first binds to a branched stem loop (stem-loop IIA/IIB/IIC) at nt 103-106 and 123-131 of a minimal 233 nt RRE (30) and, at higher Rev concentrations, also binds to stem-loop III/IV, which may allow structural changes and binding of cellular factors. Rev multimerization on the RNA is believed to be required (41). The equine infectious anemia virus (EIAV) RRE and the HERV-K RcRE also may bind their respective regulatory proteins in two distinct regions (15, 38). Use of Rev peptides and an oligomer derived from HIV stem-loop IIB have indicated considerable flexibility of the RRE (42). Purified Rem protein has been difficult to produce, but preliminary data using filter-binding assays indicate that Rem SP binds with comparable affinities to both the full-length and 1-348 RmREs (A. Chadee, unpublished results). By analogy to the RRE, secondary structural changes at the 3' end of the mutant RmRE may hamper necessary conformational changes that allow recruitment of cell proteins. The requirement for flexibility of the response element structure is supported by experiments showing that engineering of a completely base-paired stem (10 bp) at the ends of the RmRE abolished Rem responsiveness (J. Mertz, unpublished observations).

Considerable variability in size and location has been reported for the response elements that bind the Rev-like regulatory proteins of complex retroviruses. The EIAV RRE has been reported to be 555 nt near the 5' end of the envelope gene (43), but a minimal 57 nt sequence that spans the exonic splicing enhancer can function in a nuclear

export assay (44). The minimal HIV RRE has been reported to be 234 nt (29) and is localized to the envelope gene (29, 45), whereas the fully active element appears to be 351 nt (45). The HTLV-1 RxRE is 254 nt and spans the U3 and R regions of the 3' LTR (13, 14, 46). The full-length Rec-responsive element (RcRE) was reported to be 416 to 429 nt and also localized to the U3/R border (15, 16). However, like the RmRE, a 374 nt RcRE deleted from the 3' end had partial activity (47).

Mullner et al. reported that deletion of a 490 bp fragment spanning the MMTV envelope-3' LTR junction abolished Gag expression from an MMTV molecular clone (11), which is consistent with our original report (1). This fragment and various deleted versions then were inserted into a heterologous HIV-based reporter vector for analysis of HIV Gag expression in cat kidney cells, which lack endogenous MMTV proviruses (48). Although the response was not quantitated, a minimal element of 279 nt was shown to be Rem-responsive. Our data indicate that the 1-348 RmRE is minimally functional, and the 3' end of this sequence extends beyond their minimal element. Deletion of 30 bases at the 5' end of our RmRE retained considerable activity (Figs. 2 and 5); this deletion includes stem I of their sequence, which also was not required in their mapping experiments (11). Sequences spanning their proposed stem IIA, IIB, and IIC were essential for activity in their assays as well as ours. Our data indicate that part of region I and all of regions II, III, and IV are required for full RmRE function (Fig.A4). Differences with our results and those of Mullner et al. (11) could be due to the context of the sequences in the different reporter plasmids or use of different cell lines.

Our results also differ from the extended hairpin model of Mullner et al. (11) using co-variation analysis. A transition from C to U at nt 90 was accompanied in four

MMTV isolates by a G to A change at nt 95, positions that are predicted to be base paired in the model constrained by the RNase protection data (Fig.A4). Further, a G to A change at nt 136 of four additional MMTV isolates was accompanied by a C to U change at nt 166, positions that are predicted to be base paired in the unconstrained structure (Fig. S2). These data may indicate that the RmRE undergoes structural rearrangements between the unconstrained and constrained structures similar to reports with the HIV RRE (30). However, we should note that co-variation analysis is limited by the availability and accuracy of sequence information from different infectious MMTVs as well as the multiple constraints on the RmRE. The predicted RmRE overlaps with coding regions for the envelope and superantigen proteins as well as cis-acting sequences needed for integration and transcription. Nevertheless, co-variation analysis did not support the extended hairpin model (11).

Lentivirus response elements appear to map to SU or the SU/TM junctions that will be spliced out of many subgenomic mRNAs, yet a number of complex retroviruses have response elements that localize to genomic regions present in all viral mRNAs. Therefore, the idea that Rev-like regulatory proteins have functions other than nuclear export of intron-containing viral RNAs seems likely (49-53). Experiments using two different betaretroviruses, Jaagsiekte sheep retrovirus and MMTV, suggest that Rev-like regulatory proteins that include the envelope signal peptide have a post-export activity (24, 54). Indeed, Rem activity in the pHM*Rluc* assay appears to function after nuclear export (24), although Rem also has been shown to affect export of the unspliced MMTV RNA (1, 2, 11). Interestingly, experiments by Rizvi et al. suggest that RNAs transcribed from MMTV-based Gag-Pol expression vectors could be exported to the cytoplasm in the

absence of the RmRE (55). However, these RNAs required a Mason-Pfizer monkey virus (MPMV) CTE generally found in simple retroviruses for efficient Gag and Pol protein expression, but not for RNA export. These results are consistent with the interpretation that the MPMV CTE substitutes for the post-export function of Rem.

The boundaries of the MMTV RmRE also were tested in two different cell lines using a more highly active form of Rem, P71L (Fig.A5). Unlike the originally reported sequence (1), leucine is found in position 71 of the signal peptide in both the RIII and C3H MMTV exogenous viruses (56 ,57). The presence of proline may impair Rem cleavage to SP by signal peptidase. A GFP-SP fusion, which correlates with the higher activity in reporter assays, was observed by Western blotting (Fig.A5-A). Although higher reporter activity was obtained with P71L, no differences in the limits of the RmRE were obtained using either XC rat fibroblasts or human Jurkat T cells. Such results indicate that the cellular machinery that interacts with the RmRE is conserved between rodents and humans.

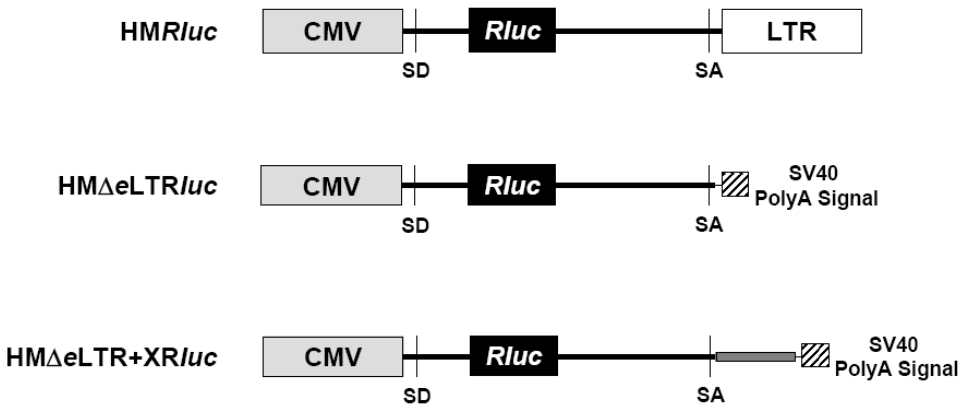

Figure A1: Diagram of reporter vectors used to map the MMTV RmRE

The CMV promoter, the *Renilla* luciferase gene, the MMTV 3' LTR, and SV40 polyadenylation sequences are shown as gray, black, white, and hatched boxes, respectively. Splice donor (SD) and splice acceptor (SA) sites also are shown within MMTV-derived sequences. The smaller black box in pHMΔeLTR+X*Rluc* represents the "X" number of MMTV sequences inserted in the reporter vectors to characterize the limits of the RmRE.

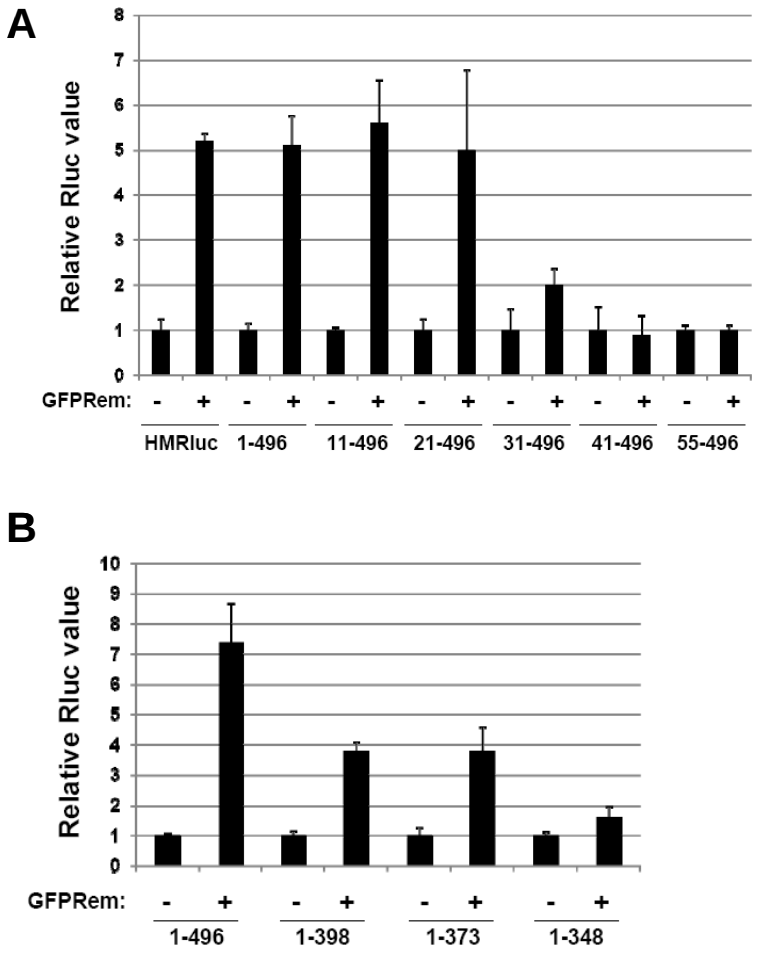

Figure A2: 5' and 3' deletion mutants define the limits of the RmRE

A, Transient transfections of rat XC fibroblasts using reporter constructs containing different 5' deletions of the MMTV RmRE. Each construct was co-transfected using DMRIE-C in the presence or absence of GFP-tagged Rem. Transfections were performed in triplicate, and the averages and standard deviations are shown. Vector transfections in the absence of Rem were assigned a relative Renilla luciferase (Rluc) value of 1, and induction by Rem-GFP co-transfection is shown compared to this normalized value. **B**, Transient transfections of rat XC fibroblasts using reporter constructs containing different 3' deletions of the MMTV RmRE. Experimental values are reported as in panel A. Unnormalized data also have been presented (Table S1).

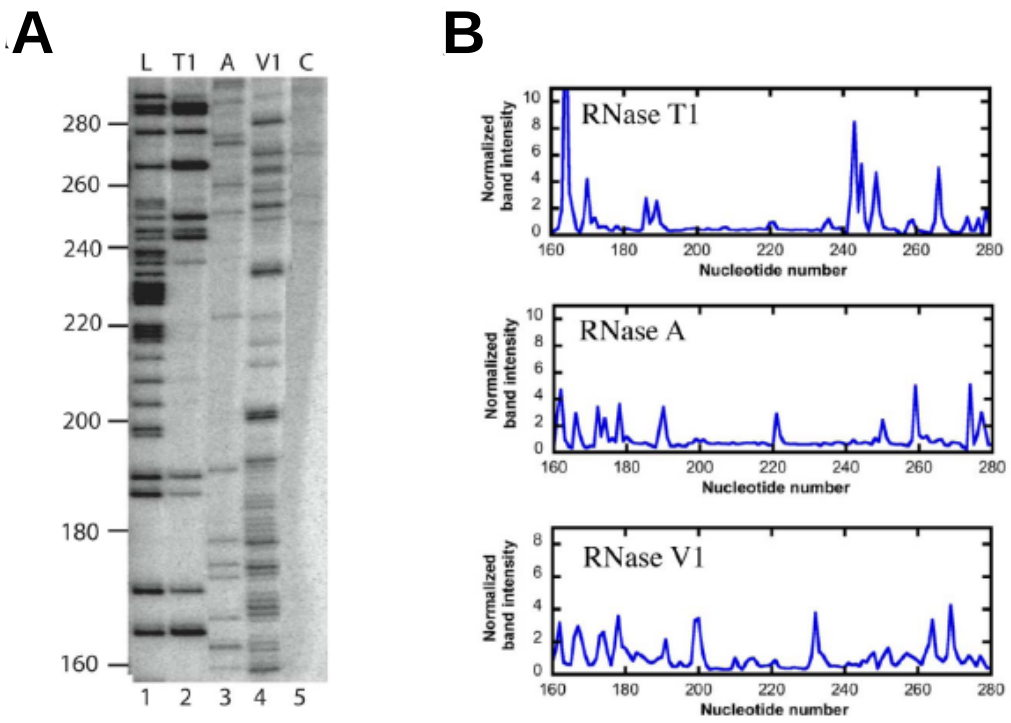

Figure A3: Nuclease mapping of the MMTV RmRE

A, Gel image of nucleotides 160-280 of the RmRE after cleavage of end-labeled RNA by RNases T1, A, and V1. A ladder (L) showing G residues was generated by treating the RNA with RNase T1 under denaturing conditions. In the control lane (C), the RNA was treated under the same conditions, but without the addition of any RNases. Lane V1 shows cleavage by RNase V1 and suggests base pairing in the secondary structure. In lane A, RNA was treated with RNase A which specifically cleaves single-stranded C or U residues and, in lane T1, RNA was treated with RNase T1, which preferentially cleaves single-stranded G residues. **B,** Quantitation of digestions with each RNase within the region of the RmRE shown in panel A. Band intensities are shown as normalized values (see Experimental Procedures) and reflect averages from two to six independent determinations. Analogous quantitation of the entire RmRE sequence using the software SAFA is shown in supplementary Figs. S1-S3.

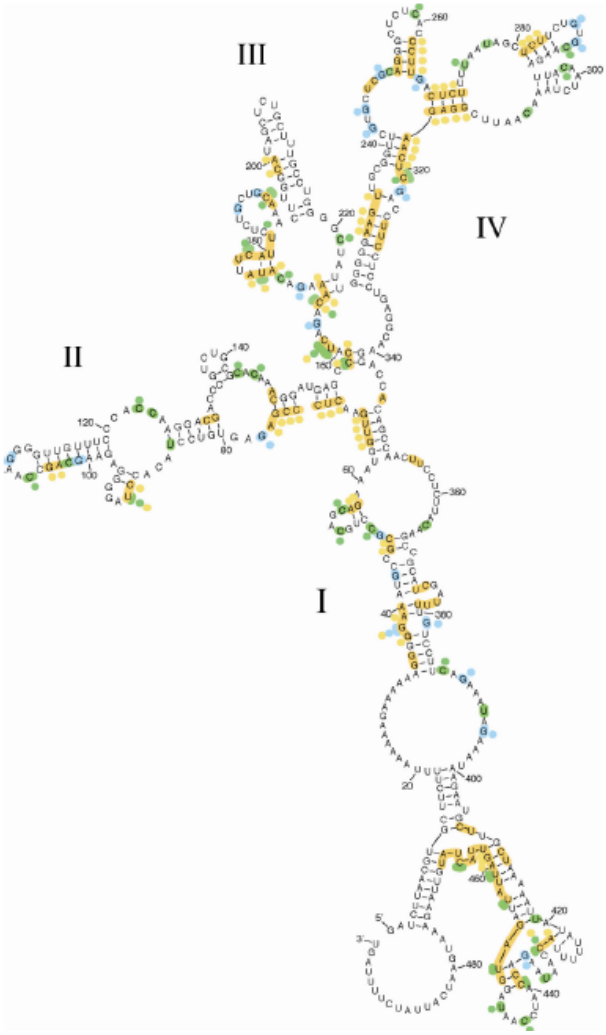

Figure A4: Secondary structure prediction with experimental constraints for the MMTV RmRE

Cleavage sites with RNase V1, A, and T1 are highlighted in yellow, green and blue, respectively. Those sites indicated by colored circles were used as constraints in a web-based version of Mfold (28) to generate the secondary structure prediction shown. The shaded regions that are not indicated with circles represent nucleotides that reacted with the indicated RNase, but did not give a level of cleavage sufficient to warrant inclusion as a constraint in the prediction. Nucleotide cleavage data represent an average of two to six independent experiments for all three nucleases with the exception of residues 315-370 for RNase V1 and residues 314-340 for RNase A (one determination).

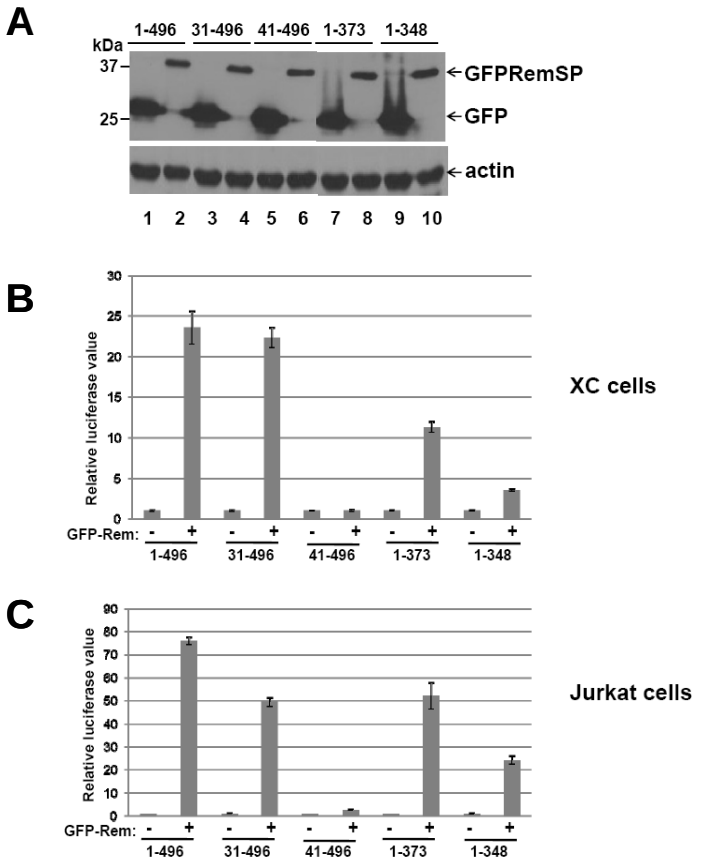

Figure A5: Activities of 5' and 3' RmRE deletion mutants with RemP71L in XC rat fibroblasts and Jurkat human T cells

A, Western blotting of cell extracts transfected with GFP-tagged RemP71L. Western blots were incubated with GFP-specific antibody (top panel) and then stripped and incubated with actin-specific antibody (bottom panel) as indicated. The positions of EGFP and GFP-tagged RemSP are shown on the right by arrows. A small amount of full-length GFP-RemP71L could be observed by overexposure of the blot (not shown). The positions of molecular mass markers are shown on the left. **B,** Transfections of the 1-496 RmRE reporter vector and deletion mutants in XC rat fibroblast cells using GFP-tagged RemP71L. **C**, Transfections of the 1-496 RmRE reporter vector and deletion mutants in Jurkat human T cells using GFP-tagged RemP71L. The averages and standard deviations of each reporter vector (full-length 1-496 or mutants) transfected in the absence of co-transfected Rem were assigned a value of 1. Levels of induction in the presence of co-transfected Rem expression vector are given compared to this value after normalization for transfection efficiency using a plasmid control vector lacking the RmRE (pGL3).

```
C3H (Varmus)  GAUCUUAACGUGCUUCUUUUAAAAAAGAAAAAAGGGGGAAAUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACCUAGGGGAGAAGCAGCCAAGGGGUUGUUUCC 120
RIII/Sa2 (P)  GAUCUUAACGUGCUUCUUUUAAAAAAGAAAAAAGGGGGAAAUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACUUAGGGGAGAAGCAGCCAAGGGGUUGUUUCC 120
C3H (Ross)    GAUCUUAACGUGCUUCUUUUAAAAAAGAAAAAAGGGGGAAAUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACCUAGGGGAGAAGCAGCCAAGGGGUUGUUUCC 120
C3H/HeJ       GAUCUUAACGUGCUUCUUUUAAAAAAGAAAAAAGGGGGAAAUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACUUAGGAGAGAAGCAGCCAAGGGGUUGUUUCC 120
BR6           GAUCUUAACGUGCUUCUUUUAAAAAAGAAAAAAGGGGGAAAUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACUUAGGGGAGAAGCAGCCAAGGGGUUGUUUCC 120
GR            ----------UGCUUCUUUUAAAAAAGAAAAAAGGGGGAAAUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACUUAGGGGAGAAGCAGCCAAGGGGUUGUUUCC 120
BALB14        GAUCUUAACGUGCUUCUUUUAAAAAAGAAAAAGGGGGGAAAUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACUUAGGGG----------------UUGUUUCC 120
TES14         ----------------------------------------AUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACUUAGGGGAGAAGCAGCCGAGGGUUGUUUCC 120
CS            ----------------------------------------AUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACUUAGGGGAGAAGCAGCCAAGGGGUUGUUUCC 120
TES2          ----------------------------------------AUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACUUAGGGGAGAAGCAGCCAAGGGGUUGUUUCC 120
FM            ----------------------------------------AUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACUUAGGGGAGAAGCAGCCAAGGGGUUGUUUCC 120
SW            ----------------------------------------AUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACUUAGGAGAGAAGCAGCCAAGGGGUUGUUUCC 120
RIII/Sa1(S)   ----------------------------------------AUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACUUAGGGGAGAAGCAGCCAAGGGGUUGUUUCC 120
RIII/Sa2(S)   ----------------------------------------AUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGCCUACACUUAGGGGAGAAGCAGCCAAGGGGUUGUUUCC 120
JYG           ----------------------------------------AUGCCGCGCCUGCAGCAGAAAUGGUUGAACUCCCGAGAGUGUCCUACACUUAGGAGAGAAGCAGCCAAGGGGUUGUUUCC 120
BALB2         ------------------------------------------------------------------------UCCCGAGAGUGUCCUACACUUAGGAGAGAAGCAGCCAAGGGGUUGUUUCC 120
                                                                                        *********** ******* **** *          ********

C3H (Varmus)  CACCAAGGACGACCCGUCUGCGCACAAACGGAUGAGCCCAUCAGACAAAGACAUAUUCAUUCUCUGCUGCAAACUUGGCAUAGCUCUGCUUUGCCUGGGGCUAUUGGGGGAAGUUGCGGU 240
RIII/Sa2 (P)  CACCAAGGACGACCCAUCUGCGCACACACGGAUGAGCCCAUCAGAUAAAGACAUACUCAUUCUCUGCUGCAAACUUGGCAUAGCUCUACUUUGCCUGGGGCUAUUGGGGGAAGUUGCGGU 240
C3H(Ross)     CACCAAGGACGACCCGUCUGCGCACAAACGGAUGAGCCCAUCAGACAAAGACAUAUUCAUUCUCUGCUGCAAACUUGGCAUAGCUCUGCUUUGCUGGGGGCCAUUGGGGGAAGUUGCGGU 240
C3H/HeJ       CACCAAGGACGACCCGUCUGCGUGCACGCGGAUGAGCCCAUCAGACAAAGACAUACUCAUUCUCUGCUGCAAACUUGGCAUAGCUCUGCUUUGCCUGGGGCUAUUGGGGGAAGUUGCGGU 240
BR6           CACCAAGGACGACCCGUCUGCGCACAAACGGAUGAGCCCAUCAGACAAAGACAUACUCAUUCUCUGCUGCAAACUUGGCAUAGCUCUGCUUUGCCUGGGGCUAUUGGGGGAAGUUGCGGU 240
GR            CACCAAGGACGACCCGUCUGCGCACAAACGGAUGAGCCCAUCAGACAAAGACAUACUCAUUCUCUGCUGCAAACUUGGCAUAGCUCUGCUUUGCCUAGGGCUAUUGGGGGAAGUUGCGGU 240
BALB14        CACCAAGGACGACCCGUCUGCGCACAAACGGAUGAGCCCAUCAGACAAAGACAUACUCAUUCUCUGCUGCAAACUUGGCAUAGCUCUACUUUGCCUGGGGCUAUUGGGGGAAUUUGCGGU 240
TES14         CACCAAGGACGACCCAUCUGCGCACAAACGGAUGAGCCCAUCAGACAAAGACAUACUCAUUCUCUGCUGCAAACUUGGCAUAGCUCUGCUUUGCCUGGGGCUAUUGGGGGAAGUUGCGGU 240
CS            CACCAAGGACGACCCAUCUGCGCACACACGGAUGAGCCCAUCAGAUAAAGACAUACUCAUUCUCUGCUGCAAACUUGGCAUAGCUCUACUUUGCCUGGGACUAUUGGGGGAAGUUGCGGU 240
TES2          CACCAAGGACGACCCAUCUGCGCACAAACGGAUGAGCCCAUCAGAUAAAGACAUACUCAUUCUCUGCUGCAAACUUGGCAUAGCUCUGCUUUGCCUGGGCUAUUGGGGGAAGUUGCGGU 240
FM            CACCAAGGACGACCCAUCUGCGCACAAACGGAUGAGCCCAUCAGACAAAGACAUACUUAUUCUCUGCUGCAAACUUGGCAUAGCUCUGCUUGCCUGGGGCUAUUGGGGGAAGUUGCGGU 240
SW            CACCAAGGACGACCCGUCUGCGUGCACGCGGAUGAGCCCAUCAGACAAAGACAUACUCAUUCUCUGCUGCAAACUUGGCAUAGCUCUGCUUUGCCUGGGGCUAUUGGGGGAAGUUGCGGU 240
RIII/Sa1 (S)  CACCAAGGACGACCCAUCUGCGCACACACGGAUGAGCCCAUCAGAUAAAGACAUACUCAUUCGCUGCUGCAAACUUGGCAUAGCUCUGCUUUGCCUGGGGCUAUUGGGGGAAGUUGCGGU 240
RIII/Sa2 (S)  CACCAAGGACGACCCAUCUGCGCACACACGGAUGAGCCCAUCAGAUAAAGACAUACUCAUUCGCUGCUGCAAACUUGGCAUAGCUCUACUUUGCCUGGGGCUAUUGGGGGAAGUUGCGGU 240
JYG           CACCAAGGACGACCCGUCUGCGUGCACGCGGAUGAGCCCAACAGACAAAGACCUGCUCAUUCUCUGCUGCAAACUUGGCAUAGCUCUGCUUUGCCUGGGGCUACUGGGGGAAGUUGCGGU 240
BALB2         CACCAAGGACGACCCGUCUGCGUGCACGCGGAUGAGCCCAUCAGACAAAGACAUACUCAUUCUCUGCUGCAAACUUGGCAUAGCUCUGCUUUGCCUGGGGCUAUUGGGGGAAGUUGCGGU 240
              *************** ******  **  ************ **** ****** *  * **** ************************* ** ***   ** * * ******** *******

C3H (Varmus)  UCGUGCUCGCAGGGCUCUCACCCUUGACUCUUU---UAAUAGCUCUUCUGUGCAAGAUUACAAUCUAAACAAUUCGGAGAACUCGACCUUCCUCCUGAGGCAAGGACCACAGCCAACUUC 357
RIII/Sa2 (P)  UCGUGCUCGCAGGGCUCUCACCCUUGACUCUUU---GAAUAGCUCUUCUGUGCAAGAUUACAAUCUAAACGAUUCGGAGAACUCAGCCUUCCUCCUGGGGCAAGGACCACAGCCAACUUC 357
C3H (Ross)    UCGUGCUCGCAGGGCUCUCACCCUUGACUCUUU---UAAUAGCUCUUCUGUGCAAGAUUACAAUCUAAACGAUUCGGAGAACUCGACCUUCCUCCUGAGGCAAGGACCACAGCCAACUUC 357
C3H/HeJ       UCGUGCUCGCAGGGCUCUCACCCUUGACUCUUU---UAAUAACUCUUCUGUGCAAGAUUACAAUCUAAACGAUUCGGAGAACUCGACCUUCCUCCUGGGGCAAGGACCACAGCCAACUUC 357
BR6           UCGUGCUCGCAGGGCUCUCACCCUUGACUCUUU---UAAUAGCUCUUCUGUGCAAGAUUACAAUCUAAACGAUUCGGAGAACUCGACCUUCCUCCUGAGGCAAGGACCACAGCCAACUUC 357
GR            UCGUGCUCGCAGGGCUCUCACCCUUGAUUCUUUCAAUAAUAACUCUUCUGUGCAAGAUUACAAUCUAAACGAUUCGGAGAACUCGACCUUCCUCUUGGGGCAAGGACCACAGCCAACUUC 360
BALB14        UCGUGCUCGCAGGGCUCUCACCCUUGACUCUUU---GAAUAGCUCUUCUGUGCAAGAUUACAAUCUAAACAAUUCGGAGAACUCGACCUUCCUCCUGGGGCAAAGACCACAGCCAACUUC 357
TES14         UCGUGCUCGCAGGGCUCUCACCCUUGACUCUUU---UAAUAGCUCUUCUGUGCAAGAUUACAAUCUAAACAAUUCAGAGAACUCGACCUUCCUCCUGAGGCAAGGACCACAGCCAACUUC 357
CS            UCGUGCUCGCAGGGCUCUCACCCUUGACUCUUU---GAAUAGCUCUUCUGUGCAAGAUUACAAUCUAAACAAUUCGGAGAACUCGACCUUCCUCCUGGGGCAAGGACCACAGCCAACUUC 357
TES2          UCGUGCUCGCAGGGCUCUCACCUUUGACUCUUU---GAAUAGCUCUUCUGUGCAAGAUUACAAUCUAAACAAUUCGGAGAACUCGACCUUCCUCCUGAGGCAAGGACCACAGCCAACUUC 357
FM            UCGUGCUCGCAGGGCUCUCACCCUUGACUCUCU---UAAUAGCUCUUCUGUGCAAGAUUACAAUCUAAACGAUUCGGAGAACUCGACCUUCCUCCUGGGGCAAGGACCACAGCCAACUUC 357
SW            UCGUGCUUGCAGGGCUCUCACCCUUGAUUCUUU---UAAUAACUCUUCUGUGCAAGAUUACAAUCUAAGCAAUUCGGAGAACUCGACCUUCCUCCUGGGGCAAGGACCACAGCCAACUUC 357
RIII/Sa1 (S)  UCGUGCUCGCAGGGCUCUCACCAUCGACUCUUU---GAAUAGCUCUUCUGUGCAAGAUUACAAUCUAAGCAAUUCGGAGAACUCGACCUUCCUCCUGGGGCAAGGACCACAGCCAACUUC 357
RIII/Sa2 (S)  UCGUGCUCGCAGGGCUCUCACCAUCGACUCUUU---GAAUAGCUCUUCUGUGCAAGAUUACAAUCUAAGCAAUUCGGAGAACUCGACCUUCCUCCUGGGGCAAGGACCACAGCCAACUUC 357
JYG           UCGUGCUCGCAGGGCUCUCUCUCUCACAAUGCUUU---UAAUAGUUCUUCUGUGCAAGAUUACAAUCUAAACAAUUCGGAGAACUCGACCUUCCUCCUGGGGCAAGAACCACAGCCAACUUC 357
BALB2         UCGUGCUCGCAGGGCUCUCACCCUUGAUUCUUU---UAAUAACUCUUCUGUGCAAGAUUACAAUCUAAACGAUUCGGAGAACUCGACCUUCCUCCUGGGGCAAGGACCACAGCCAACUUC 357
              ******* *********** *     *  ** *    ****  ************************** * **** ******** ******** ** *****  ***************

C3H (Varmus)  CUCUUACAAGCCGCAUCGA-UUUUGUCCUUCAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAUUAUAUUUUUACCAAUAAGACCAAUCCAAUAGGUAGAUUAUUAGUUACUAUGUUAAGAA 476
RIII/Sa2 (P)  CUCUUACAAGCCACACCGA-CUUUGUCCUUCAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAUUAUAUUUUUACCAAUAAGACCAAUCCAAUAGGUAGAUUAUUAGUCACUAUGUUAAGAA 476
C3H (Ross)    CUCUUACAAGCCGCAUCGA-UUUUGUCCUUCAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAUUAUAUUUUUACCAAUAAGACCAAUCCAAUAGGUAGAUUAUUAGUUACUAUGUUAAGAA 476
C3H/HeJ       CUCUUACAAGCCGCAUCGA-AUUUGUCCUUUAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAGUAUAUUUUUACCAAUAAGACCAAUCCAAUAGGUCGAUUAUUAGUCACUAUGUUAAGAA 479
BR6           CUCUUACAAGCCGCAUCAA-CCUUGUCCUUCAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAUUAUAUUUUUACCAAUAAGACCAAUCCAAUAGGUAGAUUAUUAGUUACUAUGUUAAGAA 476
GR            CUCUUACAAGCCGCAUCGA-CUUUGUCCUUCAAAAAUAGAAAUAAGAAUGCUUGCUAAAAAUUAUGUUUUUACCAAUAAGACCAAUCCAAUAGGUCGAUUAUUAGUUACUAUGUUAAGAA 476
BALB14        CUCUUACAAGCCGCAUCGACUUUUGUCCUUCAGAA-UAGAAAUAAGAAUGCUUGCUAAAAAUUAUAUUUUUACCAAUAAGACCAAUCCAAUAGGUAGAUUAUUAGUUACUAUGUUAAGAA 476
TES14         CUCUUACAAGCCGCAUCGA-UUUUGUCCUUUAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAUUACAUUUUUACCAAUAAGACCAAUCCAAUAGGUCGAUUAUUAGUCACUAUGUUAAGAA 476
CS            CUCUUACAAGCCGCAUCGA-UUUUGUCCUUUAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAUUACAUUUUUACCAAUAAGACCAAUCCAAUAGGUCGAUUAUUAGUCACUAUGUUAAGAA 476
TES2          CUCUUACAAGCCGCAUCGA-UUUUGUCCUUUAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAUUACAUUUUUACCAAUAAGACCAAUCCAAUAGGUCGAUUAUUAGUCACUAUGUUAAGAA 476
FM            CUCUUACAAGCCGCAUCGA-UUUUGUCCUUCAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAUUAUAUUUUUACCAAUAAGACCAAUCCAAUAGGUCGAUUGUUAAUUACUAUGUUAAGAA 476
SW            CUCUUACAAGCCACACCGA-CUUUGUCCUUCAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAUUAUAUUUUUACCAAUAAGACCAAUCCAAUAGGUCGAUUAUUAAUCAUGAUGUUAAGAA 476
RIII/Sa1 (S)  CUCUUACAAGCCGCAUCGA-UUUUGUCCUUCAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAAUAUAUUUUUACCAAUAAGACCAAUCCAAUAGGUAGAUUAUUAGUCACUAUGUUAAGAA 476
RIII/Sa2 (S)  CUCUUACAAGCCGCAUCGA-UUUUGUCCUUCAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAUUAUACUUUUACCAAUAAGACCAAUCCAAUAGGUAGAUUACUAAUUACUAUGUUAAGAA 476
JYG           CUCUUACAAGCCACACCGA-UUCUGUCCUUUAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAUUAUACUUUUACCAAUAAGACUAAUCCAAUAGGUAGAUUACUAAUUACUAUGUUAAGAA 476
BALB2         CUCUUACAAGCCACACCGA-CUUUGUCCUUCAGAAAUAGAAAUAAGAAUGCUUGCUAAAAAUUAUGUUUUUACCAAUAAGACCAAUCCAAUAGGUCGAUUAUUAAUCAUGAUGUUAAGAA 476
              ************ ** * *    ******* * ** ************************* **   ********** **** ************ ****  ** * *  **********

C3H (Varmus)  AUGAAUCAUUAUCUUUUAGU 496
RIII/Sa2 (P)  AUGAAUCAUUAUCUUUUAGU 496
C3H (Ross)    AUGAAUCAUUAUCUUUUAGU 496
C3H/HeJ       AUGAAUCUUUGUCUUUUAGC 496
BR6           AUGAAUCAUUACCUUUUAGU 496
GR            AUGAAUCAUUAUCUUUUAGU 499
BALB14        AUGAAUCGUUACCUUUUAGU 496
TES14         AUGAAUCAUUACCUUUUAGU 496
CS            AUGAAUCAUUAUCUUUUAGU 496
TES2          AUGAAUCAUUAUCUUUUAGU 496
FM            AUGAAUCAUUACCUUUUAGU 496
SW            AUGAAUCUUUGUCUUUUAGC 496
RIII/Sa1 (S)  AUGAAUCAUUAUCUUUUAGU 496
RIII/Sa2 (S)  AUGAAUCAUUAUCUUUUAGU 496
JYG           AUGAAUCUCUUUCUUUUAGU 496
BALB2-MMTV    AUGAAUCUUUGUCUUUUAGC 496
              *******  *  *******
```

Figure A6: Co-variation analysis of sequences from different infectious MMTV strains

Available nucleotide sequences from milk-borne MMTV strains or endogenous proviruses that also are transmitted through milk (*Mtv2* (aka GR-MMTV) and *Mtv1*) were aligned. Asterisks indicate the bases that are completely conserved. The dashes at the beginning of the sequence (through nt 40) as well as nucleotides 97 through 112 of BALB14-MMTV indicate partial sequence availability. Sequence was obtained from the following accession numbers in PubMed: C3H (Majors), K00556; RIII/Sa2 (Popko), DQ767968; C3H (Ross), AF228552; C3H/HeJ, AF228551; Mtv1, AF228550; BR6, M15122; BALB14; Susan Ross, unpublished data and U71270; BALB2, Susan Ross, unpublished data and U71271; JYG, L11749; TES14, D38639; CS, D49536; TES2, D45409; FM, D26359; SW, X65340; GR, V01175; RIII/Sa1 (Sarkar), AF136898; RIII/Sa2 (Sarkar), AF136899. Non-conserved bases were examined for base pairing in the constrained and unconstrained RmRE structures (Fig.A4 and S2, respectively) as well as the extended hairpin model of Mullner et al. (11). Co-variations in predicted base-paired regions are noted between positions 90 and 95 (bold) and between positions 136 and 166 (bold and underlined).

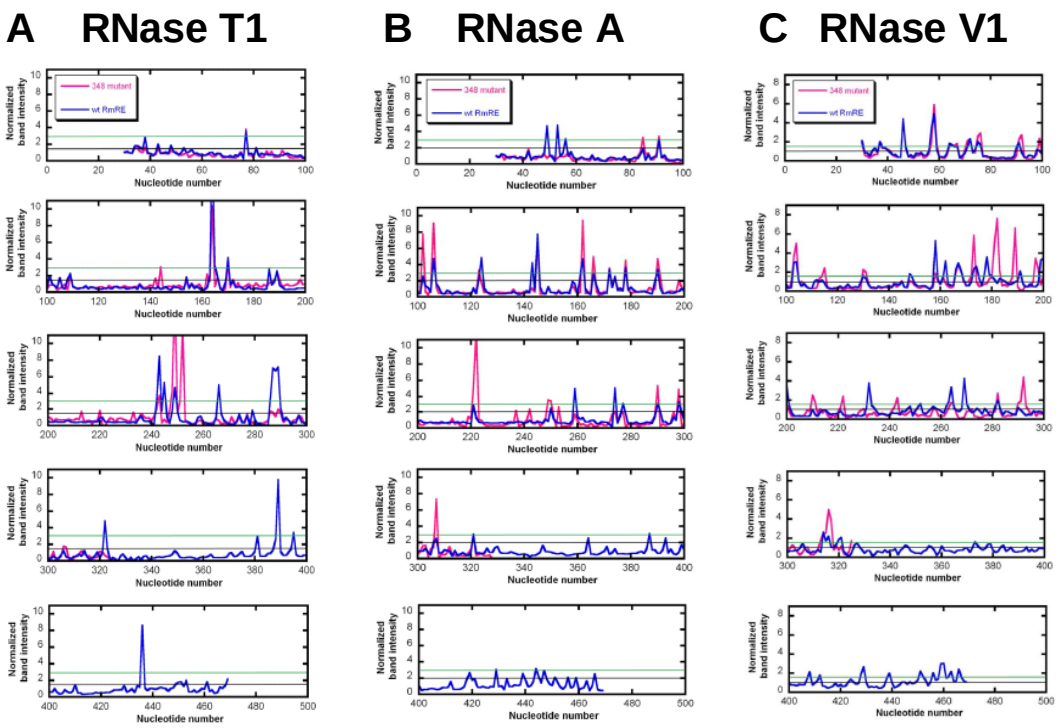

Figure S1. RNase mapping of the wild-type RmRE (blue) and the 1-348 mutant (pink) using RNase T1, RNase A and RNase T1

A,Normalized band intensity at each nucleotide position allows for a quantitative measure of residues cleaved.Peaks above the normalized intensity of 1.5 (black line) are mapped as cleaved by RNase T1. A second threshold was set to include residues above the normalized intensity 3 (green line), which were used as experimental constraints in the structure prediction. Regions of differences between the wild-type and 1-348 deletion mutant RmREs were observed beyond nt 140, with large scale changes occurring beyond nt 240.**B**, Mapping with RNase A. For the wild-type RmRE, band intensities over 2 (black line) are mapped as cleaved by RNase A. A second threshold used in constraints (green line) includes peaks over a normalized intensity of 3. Differences observed between the wild-type and 348 mutant RmREs are close to the 5' end, with the majority of larger differences observed beyond nt 220.**C**, Mapping using RNase V1.bands with normalized intensities over 1 (black line) are mapped as cleaved by RNase V1. Band intensities over 1.5 (green line) were used as the constrained limit. This limit is lower than the other RNases to reflect smaller values of "hits". The smaller values are due in part to the larger number of reactive nucleotides (e.g., see raw data in Figure 3), effectively raising the average and resulting in a smaller deviation from the average reactivity. In addition, a decreased dynamic range of reactivity was observed for this probe. Large scale differences in base-paired regions between the wild-type and 348 mutant RmREs are observed beyond nucleotide 158, suggesting structural rearrangements on the 3' side of the response element.

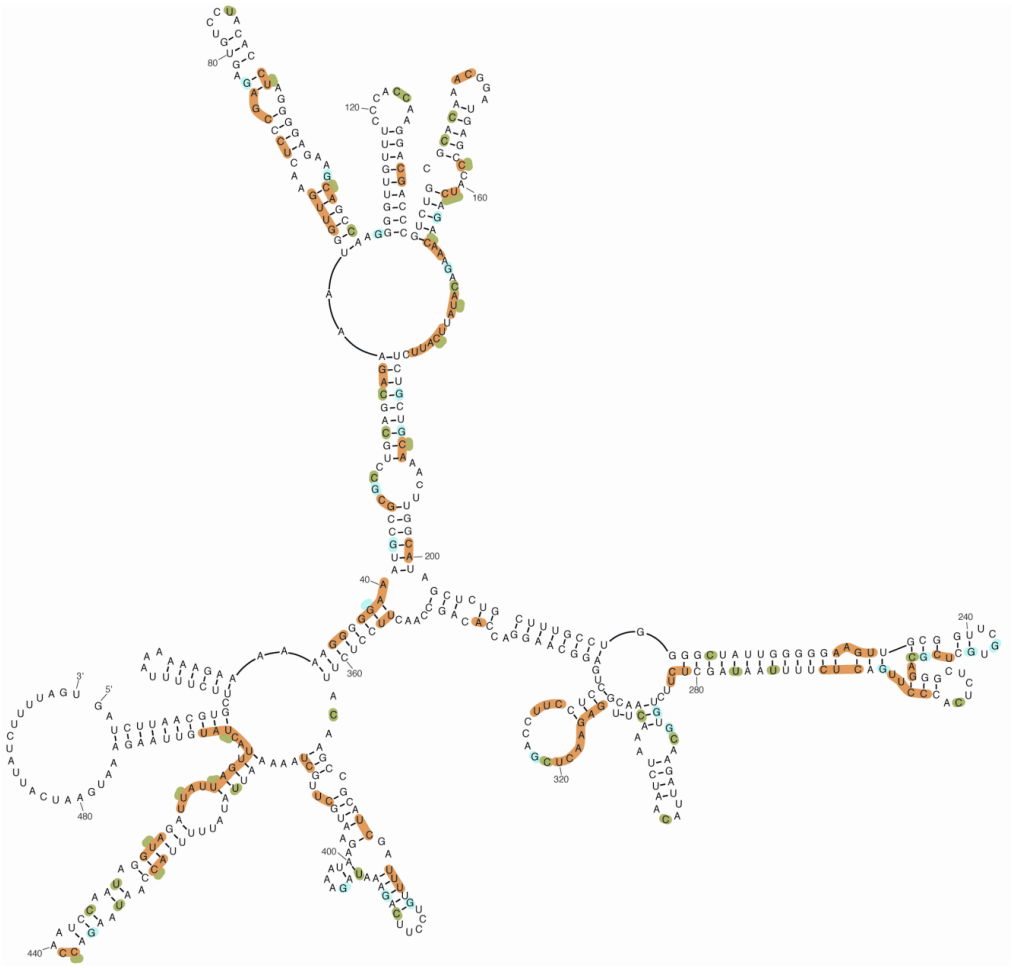

Figure S2. Predicted secondary structure of the MMTV RmRE in the absence of experimental constraints

The lowest free energy structure obtained using Mfold is shown. Results of footprinting results are also indicated (superimposed on the structure). Cleavage sites of RNase V1, A, and T1 are highlighted in brown, green and blue, respectively.

Appendix References

1. Mertz, J. A., Simper, M. S., Lozano, M. M., Payne, S. M., and Dudley, J. P. (2005) J. Virol. 79, 14737-14747

2. Indik, S., Gunzburg, W. H., Salmons, B., and Rouault, F. (2005) Virology 337, 1-6

3. Bergman, A. C., Bjornberg, O., Nord, J., Nyman, P. O., and Rosengren, A. M. (1994) Virology 204, 420-424

4. Payne, S. L. and Elder, J. H. (2001) Curr. Protein Pept. Sci. 2, 381-388

5. Golovkina, T. V., Chervonsky, A., Dudley, J. P., and Ross, S. R. (1992) Cell 69, 637-645

6. Held, W., Waanders, G. A., Shakhov, A. N., Scarpellino, L., Acha-Orbea, H., and MacDonald, H. R. (1993) Cell 74, 529-540

7. Bhadra, S., Lozano, M. M., Payne, S. M., and Dudley, J. P. (2006) PLoS. Pathog. 2, e128

8. Malim, M. H. and Emerman, M. (2008) Cell Host. Microbe. 3, 388-398

9. Neville, M., Stutz, F., Lee, L., Davis, L. I., and Rosbash, M. (1997) Curr. Biol. 7, 767-775

10. Gruter, P., Tabernero, C., von, K. C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B. K., and Izaurralde, E. (1998) Mol. Cell. 1, 649-659

11. Mullner, M., Salmons, B., Gunzburg, W. H., and Indik, S. (2008) Nucleic Acids Res. 36, 6284-6294

12. Ahmed, Y. F., Hanly, S. M., Malim, M. H., Cullen, B. R., and Greene, W. C. (1990) Genes Dev. 4, 1014-1022

13. Toyoshima, H., Itoh, M., Inoue, J., Seiki, M., Takaku, F., and Yoshida, M. (1990) J. Virol. 64, 2825-2832

14. Hanly, S. M., Rimsky, L. T., Malim, M. H., Kim, J. H., Hauber, J., Duc, D. M., Le, S. Y., Maizel, J. V., Cullen, B. R., and Greene, W. C. (1989) Genes Dev. 3, 1534-1544

15. Yang, J., Bogerd, H., Le, S. Y., and Cullen, B. R. (2000) RNA. 6, 1551-1564

16. Magin-Lachmann, C., Hahn, S., Strobel, H., Held, U., Lower, J., and Lower, R. (2001) J. Virol. 75, 10359-10371

17. Groom, H. C., Anderson, E. C., Dangerfield, J. A., and Lever, A. M. (2009) J. Gen. Virol. 90, 1141-1147

18. Holland, S. M., Ahmad, N., Maitra, R. K., Wingfield, P., and Venkatesan, S. (1990) J. Virol. 64, 5966-5975

19. Askjaer, P., Jensen, T. H., Nilsson, J., Englmeier, L., and Kjems, J. (1998) J. Biol. Chem. 273, 33414-33422

20. Mertz, J. A., Mustafa, F., Meyers, S., and Dudley, J. P. (2001) 75, 2174-2184

21. Donis-Keller, H., Maxam, A. M., and Gilbert, W. (1977) Nucleic Acids Res. 4, 2527-2538

22. England, T. E. and Uhlenbeck, O. C. (1978) Nature 275, 560-561

23. Das, R., Laederach, A., Pearlman, S. M., Herschlag, D., and Altman, R. B. (2005) RNA 11, 344-354

24. Mertz, J. A., Lozano, M. M., and Dudley, J. P. (2009) Retrovirology. 6, 10

25. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P., and Ehresmann, B. (1987) Nucleic Acids Res. 15, 9109-9128

26. Kjems, J., Brown, M., Chang, D. D., and Sharp, P. A. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 683-687

27. Fuchs, G., Stein, A. J., Fu, C., Reinisch, K. M., and Wolin, S. L. (2006) Nat. Struct. Mol. Biol. 13, 1002-1009

28. Zuker, M. (2003) Nucleic Acids Res. 31, 3406-3415

29. Malim, M. H., Hauber, J., Le, S. Y., Maizel, J. V., and Cullen, B. R. (1989) Nature 338, 254-257

30. Legiewicz, M., Badorrek, C. S., Turner, K. B., Fabris, D., Hamm, T. E., Rekosh, D., Hammarskjold, M. L., and Le Grice, S. F. (2008) Proc. Natl. Acad. Sci. U. S. A. 105, 14365-14370

31. Askjaer, P. and Kjems, J. (1998) J. Biol. Chem. 273, 11463-11471

32. Dultz, E., Hildenbeutel, M., Martoglio, B., Hochman, J., Dobberstein, B., and Kapp, K. (2008) J. Biol. Chem. 283, 9966-9976

33. Mustafa, F., Bhadra, S., Johnston, D., Lozano, M., and Dudley, J. P. (2003) J. Virol. 77, 3866-3870

34. Yang, J. and Cullen, B. R. (1999) RNA 5, 1645-1655

35. Paca, R. E., Ogert, R. A., Hibbert, C. S., Izaurralde, E., and Beemon, K. L. (2000) J. Virol. 74, 9507-9514

36. Ernst, R. K., Bray, M., Rekosh, D., and Hammarskjold, M. L. (1997) Mol. Cell. Biol. 17, 135-144

37. Ernst, R. K., Bray, M., Rekosh, D., and Hammarskjold, M. L. (1997) RNA 3, 210-222

38. Lee, J. H., Culver, G., Carpenter, S., and Dobbs, D. (2008) PLoS. ONE. 3, e2272

39. Olsen, H. S., Cochrane, A. W., Dillon, P. J., Nalin, C. M., and Rosen, C. A. (1990) Genes Dev. 4, 1357-1364

40. Bogerd, H. and Greene, W. C. (1993) J. Virol. 67, 2496-2502

41. Madore, S. J., Tiley, L. S., Malim, M. H., and Cullen, B. R. (1994) Virology 202, 186-194

42. Gosser, Y., Hermann, T., Majumdar, A., Hu, W., Frederick, R., Jiang, F., Xu, W., and Patel, D. J. (2001) Nat. Struct. Biol. 8, 146-150

43. Belshan, M., Harris, M. E., Shoemaker, A. E., Hope, T. J., and Carpenter, S. (1998) J. Virol. 72, 4421-4426

44. Belshan, M., Park, G. S., Bilodeau, P., Stoltzfus, C. M., and Carpenter, S. (2000) Mol. Cell Biol. 20, 3550-3557

45. Mann, D. A., Mikaelian, I., Zemmel, R. W., Green, S. M., Lowe, A. D., Kimura, T., Singh, M., Butler, P. J., Gait, M. J., and Karn, J. (1994) J. Mol. Biol. 241, 193-207

46. Seiki, M., Inoue, J., Hidaka, M., and Yoshida, M. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 7124-7128

47. Magin, C., Hesse, J., Lower, J., and Lower, R. (2000) Virology 274, 11-16

48. Salmons, B., Groner, B., Calberg-Bacq, C. M., and Ponta, H. (1985) Virology 144, 101-114

49. Brandt, S., Blissenbach, M., Grewe, B., Konietzny, R., Grunwald, T., and Uberla, K. (2007) PLoS. Pathog. 3, e54

50. Perales, C., Carrasco, L., and Gonzalez, M. E. (2005) Biochim. Biophys. Acta 1743, 169-175

51. Arrigo, S. J. and Chen, I. S. (1991) Genes Dev. 5, 808-819

52. Lawrence, J. B., Cochrane, A. W., Johnson, C. V., Perkins, A., and Rosen, C. A. (1991) New Biol. 3, 1220-1232

53. Kusuhara, K., Anderson, M., Pettiford, S. M., and Green, P. L. (1999) J. Virol. 73, 8112-8119

54. Caporale, M., Arnaud, F., Mura, M., Golder, M., Murgia, C., and Palmarini, M. (2009) J. Virol. 83, 4591-4604

55. Rizvi, T. A., Ali, J., Phillip, P. S., Ghazawi, A., Jayanth, P., and Mustafa, F. (2009) Virology. 385, 464-472

56. Majors, J. E. and Varmus, H. E. (1983) J. Virol. 47, 495-504

57. Moore, R., Dixon, M., Smith, R., Peters, G., and Dickson, C. (1987) J. Virol. 61, 480-490

Bibliography

1. Shi, H., et al., *A misfolded form of 5S rRNA is complexed with the Ro and La autoantigens.* Rna, 1996. **2**(8): p. 769-84.

2. Tijsterman, M. and R.H. Plasterk, *Dicers at RISC; the mechanism of RNAi.* Cell, 2004. **117**(1): p. 1-3.

3. Zhang, A., et al., *The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein.* Embo J, 1998. **17**(20): p. 6061-8.

4. Tucker, B.J. and R.R. Breaker, *Riboswitches as versatile gene control elements.* Curr Opin Struct Biol, 2005. **15**(3): p. 342-8.

5. Toor, N., K.S. Keating, and A.M. Pyle, *Structural insights into RNA splicing.* Curr Opin Struct Biol, 2009. **19**(3): p. 260-6.

6. Wen, J.D., et al., *Following translation by single ribosomes one codon at a time.* Nature, 2008. **452**(7187): p. 598-603.

7. Herschlag, D., *RNA chaperones and the RNA folding problem.* J Biol Chem, 1995. **270**(36): p. 20871-4.

8. Russell, R., et al., *Small angle X-ray scattering reveals a compact intermediate in RNA folding.* Nat Struct Biol, 2000. **7**(5): p. 367-70.

9. Sclavi, B., et al., *Following the folding of RNA with time-resolved synchrotron X-ray footprinting.* Methods Enzymol, 1998. **295**: p. 379-402.

10. Shcherbakova, I. and M. Brenowitz, *Monitoring structural changes in nucleic acids with single residue spatial and millisecond time resolution by quantitative hydroxyl radical footprinting.* Nat Protoc, 2008. **3**(2): p. 288-302.

11. Zhuang, X., et al., *A single-molecule study of RNA catalysis and folding.* Science, 2000. **288**(5473): p. 2048-51.

12. Duncan, C.D. and K.M. Weeks, *SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA.* Biochemistry, 2008. **47**(33): p. 8504-13.

13. Kim, S.H. and T.R. Cech, *Three-dimensional model of the active site of the self-splicing rRNA precursor of Tetrahymena.* Proc Natl Acad Sci U S A, 1987. **84**(24): p. 8788-92.

14. Lehnert, V., et al., *New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the Tetrahymena thermophila ribozyme.* Chem Biol, 1996. **3**(12): p. 993-1009.

15. Russell, R. and D. Herschlag, *New pathways in folding of the Tetrahymena group I RNA enzyme.* J Mol Biol, 1999. **291**(5): p. 1155-67.

16. Sosnick, T.R. and T. Pan, *RNA folding: models and perspectives.* Curr Opin Struct Biol, 2003. **13**(3): p. 309-16.

17. Tinoco, I., Jr. and C. Bustamante, *How RNA folds.* J Mol Biol, 1999. **293**(2): p. 271-81.

18. Draper, D.E., D. Grilley, and A.M. Soto, *Ions and RNA folding.* Annu Rev Biophys Biomol Struct, 2005. **34**: p. 221-43.

19. Das, R., et al., *The fastest global events in RNA folding: electrostatic relaxation and tertiary collapse of the Tetrahymena ribozyme.* J Mol Biol, 2003. **332**(2): p. 311-9.

20. Zarrinkar, P.P. and J.R. Williamson, *Kinetic intermediates in RNA folding.* Science, 1994. **265**(5174): p. 918-24.

21. Caprara, M.G., G. Mohr, and A.M. Lambowitz, *A tyrosyl-tRNA synthetase protein induces tertiary folding of the group I intron catalytic core.* J Mol Biol, 1996. **257**(3): p. 512-31.

22. Akins, R.A. and A.M. Lambowitz, *A protein required for splicing group I introns in Neurospora mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof.* Cell, 1987. **50**(3): p. 331-45.

23. Schroeder, R., A. Barta, and K. Semrad, *Strategies for RNA folding and assembly.* Nat Rev Mol Cell Biol, 2004. **5**(11): p. 908-19.

24. Holley, R.W., *Structure of an alanine transfer ribonucleic acid.* Jama, 1965. **194**(8): p. 868-71.

25. Madore, E., et al., *Magnesium-dependent alternative foldings of active and inactive Escherichia coli tRNA(Glu) revealed by chemical probing.* Nucleic Acids Res, 1999. **27**(17): p. 3583-8.

26. Gartland, W.J. and N. Sueoka, *Two interconvertible forms of tryptophanyl sRNA in E. coli.* Proc Natl Acad Sci U S A, 1966. **55**(4): p. 948-56.

27. Holley, R.W., et al., *Structure of a Ribonucleic Acid.* Science, 1965. **147**: p. 1462-5.

28. Bina-Stein, M., et al., *Physical studies of denatured tRNA2Glu from Escherichia coli.* Proc Natl Acad Sci U S A, 1976. **73**(7): p. 2216-20.

29. Harris, M.E. and N.R. Pace, *Analysis of the tertiary structure of bacterial RNase P RNA.* Mol Biol Rep, 1995. **22**(2-3): p. 115-23.

30. Pace, N.R. and J.W. Brown, *Evolutionary perspective on the structure and function of ribonuclease P, a ribozyme.* J Bacteriol, 1995. **177**(8): p. 1919-28.

31. Pan, T. and T.R. Sosnick, *Intermediates and kinetic traps in the folding of a large ribozyme revealed by circular dichroism and UV absorbance spectroscopies and catalytic activity.* Nat Struct Biol, 1997. **4**(11): p. 931-8.

32. Treiber, D.K., et al., *Kinetic intermediates trapped by native interactions in RNA folding.* Science, 1998. **279**(5358): p. 1943-6.

33. Rook, M.S., D.K. Treiber, and J.R. Williamson, *Fast folding mutants of the Tetrahymena group I ribozyme reveal a rugged folding energy landscape.* J Mol Biol, 1998. **281**(4): p. 609-20.

34. Russell, R., *RNA misfolding and the action of chaperones.* Front Biosci, 2008. **13**: p. 1-20.

35. Russell, R., et al., *The paradoxical behavior of a highly structured misfolded intermediate in RNA folding.* J Mol Biol, 2006. **363**(2): p. 531-44.

36. Serganov, A., L. Huang, and D.J. Patel, *Structural insights into amino acid binding and gene control by a lysine riboswitch.* Nature, 2008. **455**(7217): p. 1263-7.

37. Schultes, E.A. and D.P. Bartel, *One sequence, two ribozymes: implications for the emergence of new ribozyme folds.* Science, 2000. **289**(5478): p. 448-52.

38. Schroeder, R., et al., *RNA folding in vivo.* Curr Opin Struct Biol, 2002. **12**(3): p. 296-300.

39. Rajkowitsch, L., et al., *RNA chaperones, RNA annealers and RNA helicases.* RNA Biol, 2007. **4**(3): p. 118-30.

40. Pan, T. and T. Sosnick, *RNA folding during transcription.* Annu Rev Biophys Biomol Struct, 2006. **35**: p. 161-75.

41. Loria, A. and T. Pan, *Domain structure of the ribozyme from eubacterial ribonuclease P.* Rna, 1996. **2**(6): p. 551-63.

42. Pan, T., et al., *Folding of a large ribozyme during transcription and the effect of the elongation factor NusA.* Proc Natl Acad Sci U S A, 1999. **96**(17): p. 9545-50.

43. Wong, T., T.R. Sosnick, and T. Pan, *Mechanistic insights on the folding of a large ribozyme during transcription.* Biochemistry, 2005. **44**(20): p. 7535-42.

44. Cech, T.R., *Self-splicing of group I introns.* Annu Rev Biochem, 1990. **59**: p. 543-68.

45. Cech, T.R., *The chemistry of self-splicing RNA and RNA enzymes.* Science, 1987. **236**(4808): p. 1532-9.

46. Cech, T.R., A.J. Zaug, and P.J. Grabowski, *In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a*

guanosine nucleotide in the excision of the intervening sequence. Cell, 1981. **27**(3 Pt 2): p. 487-96.

47. Cech, T.R., et al., *Secondary structure of the Tetrahymena ribosomal RNA intervening sequence: structural homology with fungal mitochondrial intervening sequences.* Proc Natl Acad Sci U S A, 1983. **80**(13): p. 3903-7.

48. Cech, T.R., S.H. Damberger, and R.R. Gutell, *Representation of the secondary and tertiary structure of group I introns.* Nat Struct Biol, 1994. **1**(5): p. 273-80.

49. Latham, J.A. and T.R. Cech, *Defining the inside and outside of a catalytic RNA molecule.* Science, 1989. **245**(4915): p. 276-82.

50. Zaug, A.J. and T.R. Cech, *Self-splicing RNA and an RNA enzyme in Tetrahymena.* J Protozool, 1987. **34**(4): p. 416-7.

51. Zaug, A.J., C.A. Grosshans, and T.R. Cech, *Sequence-specific endoribonuclease activity of the Tetrahymena ribozyme: enhanced cleavage of certain oligonucleotide substrates that form mismatched ribozyme-substrate complexes.* Biochemistry, 1988. **27**(25): p. 8924-31.

52. Zaug, A.J., M.D. Been, and T.R. Cech, *The Tetrahymena ribozyme acts like an RNA restriction endonuclease.* Nature, 1986. **324**(6096): p. 429-33.

53. Herschlag, D. and T.R. Cech, *Catalysis of RNA cleavage by the Tetrahymena thermophila ribozyme. 1. Kinetic description of the reaction of an RNA substrate complementary to the active site.* Biochemistry, 1990. **29**(44): p. 10159-71.

54. Pan, J. and S.A. Woodson, *Folding intermediates of a self-splicing RNA: mispairing of the catalytic core.* J Mol Biol, 1998. **280**(4): p. 597-609.

55. Pan, J., M.L. Deras, and S.A. Woodson, *Fast folding of a ribozyme by stabilizing core interactions: evidence for multiple folding pathways in RNA.* J Mol Biol, 2000. **296**(1): p. 133-44.

56. Joyce, G.F., G. van der Horst, and T. Inoue, *Catalytic activity is retained in the Tetrahymena group I intron despite removal of the large extension of element P5.* Nucleic Acids Res, 1989. **17**(19): p. 7879-89.

57. Doherty, E.A., D. Herschlag, and J.A. Doudna, *Assembly of an exceptionally stable RNA tertiary interface in a group I ribozyme.* Biochemistry, 1999. **38**(10): p. 2982-90.

58. Johnson, T.H., et al., *Structural specificity conferred by a group I RNA peripheral element.* Proc Natl Acad Sci U S A, 2005. **102**(29): p. 10176-81.

59. Naito, Y., H. Shiraishi, and T. Inoue, *P5abc of the Tetrahymena ribozyme consists of three functionally independent elements.* Rna, 1998. **4**(7): p. 837-46.

60. Engelhardt, M.A., et al., *The P5abc peripheral element facilitates preorganization of the tetrahymena group I ribozyme for catalysis.* Biochemistry, 2000. **39**(10): p. 2639-51.

61. Treiber, D.K. and J.R. Williamson, *Concerted kinetic folding of a multidomain ribozyme with a disrupted loop-receptor interaction.* J Mol Biol, 2001. **305**(1): p. 11-21.

62. Russell, R. and D. Herschlag, *Probing the folding landscape of the Tetrahymena ribozyme: commitment to form the native conformation is late in the folding pathway.* J Mol Biol, 2001. **308**(5): p. 839-51.

63. Mohr, G., et al., *A tyrosyl-tRNA synthetase can function similarly to an RNA structure in the Tetrahymena ribozyme.* Nature, 1994. **370**(6485): p. 147-50.

64. Lambowitz, A.M. and P.S. Perlman, *Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing.* Trends Biochem Sci, 1990. **15**(11): p. 440-4.

65. Herbert, C.J., et al., *The NAM2 proteins from S. cerevisiae and S. douglasii are mitochondrial leucyl-tRNA synthetases, and are involved in mRNA splicing.* Embo J, 1988. **7**(2): p. 473-83.

66. Labouesse, M., *The yeast mitochondrial leucyl-tRNA synthetase is a splicing factor for the excision of several group I introns.* Mol Gen Genet, 1990. **224**(2): p. 209-21.

67. Weeks, K.M. and T.R. Cech, *Assembly of a ribonucleoprotein catalyst by tertiary structure capture.* Science, 1996. **271**(5247): p. 345-8.

68. Buchmueller, K.L., et al., *A collapsed non-native RNA folding state.* Nat Struct Biol, 2000. **7**(5): p. 362-6.

69. Buchmueller, K.L. and K.M. Weeks, *Near native structure in an RNA collapsed state.* Biochemistry, 2003. **42**(47): p. 13869-78.

70. Webb, A.E., et al., *Protein-dependent transition states for ribonucleoprotein assembly.* J Mol Biol, 2001. **309**(5): p. 1087-100.

71. Wallweber, G.J., et al., *Characterization of Neurospora mitochondrial group I introns reveals different CYT-18 dependent and independent splicing strategies and an alternative 3' splice site for an intron ORF.* Rna, 1997. **3**(2): p. 114-31.

72. Paukstelis, P.J., et al., *Structure of a tyrosyl-tRNA synthetase splicing factor bound to a group I intron RNA.* Nature, 2008. **451**(7174): p. 94-7.

73. Saldanha, R.J., et al., *Involvement of Neurospora mitochondrial tyrosyl-tRNA synthetase in RNA splicing. A new method for purifying the protein and characterization of physical and enzymatic properties pertinent to splicing.* Biochemistry, 1995. **34**(4): p. 1275-87.

74. Myers, C.A., et al., *A tyrosyl-tRNA synthetase suppresses structural defects in the two major helical domains of the group I intron catalytic core.* J Mol Biol, 1996. **262**(2): p. 87-104.

75. Mohr, G., et al., *The neurospora CYT-18 protein suppresses defects in the phage T4 td intron by stabilizing the catalytically active structure of the intron core.* Cell, 1992. **69**(3): p. 483-94.

76. Mohr, G., et al., *Function of the Neurospora crassa mitochondrial tyrosyl-tRNA synthetase in RNA splicing. Role of the idiosyncratic N-*

terminal extension and different modes of interaction with different group I introns. J Mol Biol, 2001. **307**(1): p. 75-92.

77. Paukstelis, P.J., et al., *A tyrosyl-tRNA synthetase adapted to function in group I intron splicing by acquiring a new RNA binding surface.* Mol Cell, 2005. **17**(3): p. 417-28.

78. Paukstelis, P.J. and A.M. Lambowitz, *Identification and evolution of fungal mitochondrial tyrosyl-tRNA synthetases with group I intron splicing activity.* Proc Natl Acad Sci U S A, 2008. **105**(16): p. 6010-5.

79. Russell, R., et al., *Deletion of the P5abc peripheral element accelerates early and late folding steps of the Tetrahymena group I ribozyme.* Biochemistry, 2007. **46**(17): p. 4951-61.

80. Murphy, F.L. and T.R. Cech, *An independently folding domain of RNA tertiary structure within the Tetrahymena ribozyme.* Biochemistry, 1993. **32**(20): p. 5291-300.

81. van der Horst, G., A. Christian, and T. Inoue, *Reconstitution of a group I intron self-splicing reaction with an activator RNA.* Proc Natl Acad Sci U S A, 1991. **88**(1): p. 184-8.

82. Saldanha, R., A. Ellington, and A.M. Lambowitz, *Analysis of the CYT-18 protein binding site at the junction of stacked helices in a group I intron RNA by quantitative binding assays and in vitro selection.* J Mol Biol, 1996. **261**(1): p. 23-42.

83. Majumder, A.L., et al., *Involvement of tyrosyl-tRNA synthetase in splicing of group I introns in Neurospora crassa mitochondria: biochemical and immunochemical analyses of splicing activity.* Mol Cell Biol, 1989. **9**(5): p. 2089-104.

84. Tullius, T.D. and B.A. Dombroski, *Hydroxyl radical "footprinting": high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein.* Proc Natl Acad Sci U S A, 1986. **83**(15): p. 5469-73.

85. Huang, Z. and J.W. Szostak, *A simple method for 3'-labeling of RNA.* Nucleic Acids Res, 1996. **24**(21): p. 4360-1.

86. Das, R., et al., *SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments.* Rna, 2005. **11**(3): p. 344-54.

87. Tullius, T.D. and J.A. Greenbaum, *Mapping nucleic acid structure by hydroxyl radical cleavage.* Curr Opin Chem Biol, 2005. **9**(2): p. 127-34.

88. Roth, A. and R.R. Breaker, *The Structural and Functional Diversity of Metabolite-Binding Riboswitches.* Annu Rev Biochem, 2009.

89. Serganov, A. and D.J. Patel, *Ribozymes, riboswitches and beyond: regulation of gene expression without proteins.* Nat Rev Genet, 2007. **8**(10): p. 776-90.

90. Altuvia, S., et al., *The Escherichia coli OxyS regulatory RNA represses fhlA translation by blocking ribosome binding.* Embo J, 1998. **17**(20): p. 6069-75.

91. Puglisi, J.D., S.C. Blanchard, and R. Green, *Approaching translation at atomic resolution.* Nat Struct Biol, 2000. **7**(10): p. 855-61.

92. Doudna, J.A. and T.R. Cech, *Self-assembly of a group I intron active site from its component tertiary structural domains.* Rna, 1995. **1**(1): p. 36-45.

93. Herschlag, D., *RNA chaperones and the RNA folding problem.* J. Biol. Chem., 1995. **270**(36): p. 20871-4.

94. Sigler, P.B., *An analysis of the structure of tRNA.* Annu. Rev. Biophys. Bioeng., 1975. **4**(00): p. 477-527.

95. Treiber, D.K. and J.R. Williamson, *Exposing the kinetic traps in RNA folding.* Curr. Opin. Struct. Biol., 1999. **9**(3): p. 339-45.

96. Russell, R., *RNA misfolding and the action of chaperones.* Front. Biosci., 2008. **13**: p. 1-20.

97. Kruger, K., et al., *Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena.* Cell, 1982. **31**(1): p. 147-57.

98. Cech, T.R., *Ribozymes, the first 20 years.* Biochem. Soc. Trans., 2002. **30**(Pt 6): p. 1162-6.

99. Michel, F. and E. Westhof, *Modeling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis.* J. Mol. Biol., 1990. **216**(3): p. 585-610.

100. Lehnert, V., et al., *New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the Tetrahymena thermophila ribozyme.* Chem. Biol., 1996. **3**(12): p. 993-1009.

101. Golden, B.L., et al., *A preorganized active site in the crystal structure of the Tetrahymena ribozyme.* Science, 1998. **282**(5387): p. 259-64.

102. Adams, P.L., et al., *Crystal structure of a self-splicing group I intron with both exons.* Nature, 2004. **430**(6995): p. 45-50.

103. Guo, F., A.R. Gooding, and T.R. Cech, *Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site.* Mol. Cell, 2004. **16**(3): p. 351-62.

104. Golden, B.L., H. Kim, and E. Chase, *Crystal structure of a phage Twort group I ribozyme-product complex.* Nat. Struct. Mol. Biol., 2005. **12**(1): p. 82-9.

105. Tanner, M.A. and T.R. Cech, *Activity and thermostability of the small self-splicing group I intron in the pre-tRNA(Ile) of the purple bacterium Azoarcus.* RNA, 1996. **2**: p. 74-83.

106. Cannone, J.J., et al., *The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs.* BMC Bioinformatics, 2002. **3**(1): p. 2-32.

107. van der Horst, G., A. Christian, and T. Inoue, *Reconstitution of a group I intron self-splicing reaction with an activator RNA.* Proc. Natl. Acad. Sci. U.S.A., 1991. **88**(1): p. 184-8.

108. Doherty, E.A., D. Herschlag, and J.A. Doudna, *Assembly of an exceptionally stable RNA tertiary interface in a group I ribozyme.* Biochemistry, 1999. **38**(10): p. 2982-2990.

109. Johnson, T.H., et al., *Structural specificity conferred by a group I RNA peripheral element.* Proc. Natl. Acad. Sci. U.S.A., 2005. **102**: p. 10176-10181.

110. Russell, R., et al., *The paradoxical behavior of a highly structured misfolded intermediate in RNA folding.* J. Mol. Biol., 2006. **363**: p. 531-544.

111. Russell, R. and D. Herschlag, *New pathways in folding of the Tetrahymena group I RNA enzyme.* J. Mol. Biol., 1999. **291**(5): p. 1155-67.

112. Russell, R. and D. Herschlag, *Probing the folding landscape of the Tetrahymena ribozyme: Commitment to form the native conformation is late in the folding pathway.* J. Mol. Biol., 2001. **308**(5): p. 839-51.

113. Russell, R., et al., *Deletion of the P5abc peripheral element accelerates early and late folding steps of the Tetrahymena group I ribozyme.* Biochemistry, 2007. **46**: p. 4951-61.

114. Lambowitz, A.M., et al., *Group I and Group II ribozymes as RNPs: Clues to the past and guides to the future*, in *The RNA World*, R.F. Gesteland, T.R. Cech, and J.F. Atkins, Editors. 1999, Cold Spring Harbor: New York. p. 451-485.

115. Akins, R.A. and A.M. Lambowitz, *A protein required for splicing group I introns in Neurospora mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof.* Cell, 1987. **50**(3): p. 331-45.

116. Wallweber, G.J., et al., *Characterization of Neurospora mitochondrial group I introns reveals different CYT-18 dependent and independent splicing strategies and an alternative 3' splice site for an intron ORF.* RNA, 1997. **3**(2): p. 114-31.

117. Paukstelis, P.J. and A.M. Lambowitz, *Identification and evolution of fungal mitochondrial tyrosyl-tRNA synthetases with group I intron*

splicing activity. Proc. Natl. Acad. Sci. U. S. A., 2008. **105**(16): p. 6010-5.

118. Caprara, M.G., G. Mohr, and A.M. Lambowitz, *A tyrosyl-tRNA synthetase protein induces tertiary folding of the group I intron catalytic core.* J. Mol. Biol., 1996. **257**(3): p. 512-31.

119. Myers, C.A., et al., *A tyrosyl-tRNA synthetase suppresses structural defects in the two major helical domains of the group I intron catalytic core.* J. Mol. Biol., 1996. **262**(2): p. 87-104.

120. Caprara, M.G., C.A. Myers, and A.M. Lambowitz, *Interaction of the Neurospora crassa mitochondrial tyrosyl-tRNA synthetase (CYT-18 protein) with the group I intron P4-P6 domain. Thermodynamic analysis and the role of metal ions.* J. Mol. Biol., 2001. **308**(2): p. 165-90.

121. Paukstelis, P.J., et al., *A tyrosyl-tRNA synthetase adapted to function in group I intron splicing by acquiring a new RNA binding surface.* Mol. Cell, 2005. **17**(3): p. 417-28.

122. Russell, R. and D. Herschlag, *Specificity from steric restrictions in the guanosine binding pocket of a group I ribozyme.* RNA, 1999. **5**(2): p. 158-66.

123. Saldanha, R.J., et al., *Involvement of Neurospora mitochondrial tyrosyl-tRNA synthetase in RNA splicing. A new method for purifying the protein and characterization of physical and enzymatic properties pertinent to splicing.* Biochemistry, 1995. **34**(4): p. 1275-87.

124. Webb, A.E., et al., *Protein-dependent transition states for ribonucleoprotein assembly.* J. Mol. Biol., 2001. **309**(5): p. 1087-100.

125. Rose, M.A. and K.M. Weeks, *Visualizing induced fit in early assembly of the human signal recognition particle.* Nat. Struct. Biol., 2001. **8**(6): p. 515-20.

126. Park, Y.C. and H. Bedouelle, *Dimeric tyrosyl-tRNA synthetase from Bacillus stearothermophilus unfolds through a monomeric*

intermediate. A quantitative analysis under equilibrium conditions. J. Biol. Chem., 1998. **273**(29): p. 18052-9.

127. Fersht, A., *Structure and Mechanism In Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. 1999, U.S.A.: W.H. Freeman and Company.

128. Murphy, F.L. and T.R. Cech, *GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain.* J. Mol. Biol., 1994. **236**(1): p. 49-63.

129. Cate, J.H., et al., *Crystal structure of a group I ribozyme domain: principles of RNA packing.* Science, 1996. **273**(5282): p. 1678-85.

130. Saldanha, R., A. Ellington, and A.M. Lambowitz, *Analysis of the CYT-18 protein binding site at the junction of stacked helices in a group I intron RNA by quantitative binding assays and in vitro selection.* J. Mol. Biol., 1996. **261**(1): p. 23-42.

131. Chen, X., R.R. Gutell, and A.M. Lambowitz, *Function of tyrosyl-tRNA synthetase in splicing group I introns: an induced-fit model for binding to the P4-P6 domain based on analysis of mutations at the junction of the P4-P6 stacked helices.* J. Mol. Biol., 2000. **301**(2): p. 265-83.

132. Grossberger, R., et al., *Influence of RNA structural stability on the RNA chaperone activity of the Escherichia coli protein StpA.* Nucleic Acids Res., 2005. **33**(7): p. 2280-9.

133. Mohr, G., et al., *Function of the Neurospora crassa mitochondrial tyrosyl-tRNA synthetase in RNA splicing. Role of the idiosyncratic N-terminal extension and different modes of interaction with different group I introns.* J. Mol. Biol., 2001. **307**(1): p. 75-92.

134. Bhaskaran, H. and R. Russell, *Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone.* Nature, 2007. **449**: p. 1014-1018.

135. Mohr, S., J.M. Stryker, and A.M. Lambowitz, *A DEAD-box protein functions as an ATP-dependent RNA chaperone in group I intron splicing*. Cell, 2002. **109**(6): p. 769-79.

136. Caprara, M.G., et al., *A tyrosyl-tRNA synthetase recognizes a conserved tRNA-like structural motif in the group I intron catalytic core*. Cell, 1996. **87**(6): p. 1135-45.

137. Woodson, S.A. and T.R. Cech, *Alternative secondary structures in the 5' exon affect both forward and reverse self-splicing of the Tetrahymena intervening sequence RNA.* Biochemistry, 1991. **30**(8): p. 2042-50.

138. Woodson, S.A., *Exon sequences distant from the splice junction are required for efficient self-splicing of the Tetrahymena IVS.* Nucleic Acids Res., 1992. **20**(15): p. 4027-32.

139. Duncan, C.D. and K.M. Weeks, *SHAPE analysis of long-range interactions reveals extensive and thermodynamically preferred misfolding in a fragile group I intron RNA.* Biochemistry, 2008. **47**(33): p. 8504-13.

140. Bokinsky, G., et al., *Two distinct binding modes of a protein cofactor with its target RNA.* J. Mol. Biol., 2006. **361**(4): p. 771-84.

141. Guo, Q. and A.M. Lambowitz, *A tyrosyl-tRNA synthetase binds specifically to the group I intron catalytic core.* Genes Dev, 1992. **6**(8): p. 1357-72.

142. Price, J.V., et al., *Sequence requirements for self-splicing of the Tetrahymena thermophila pre-ribosomal RNA.* Nucleic Acids Res, 1985. **13**(6): p. 1871-89.

Vita

Amanda B Chadee was born and raised on the Caribbean island of Trinidad. She is the first born to parents Ramnarine Chadee and Amryll Chadee. She attended Laskshmi Girls High school and St.George's College in Trinidad where her interest in the sciences grew. In 2000, she left Trinidad to matriculate at Claflin University in South Carolina where she actively pursued scientific research. She graduated with honors with a B.S in Biochemistry in 2003. She began her graduate work at the University of Texas at Austin in 2003 where she joined the laboratory of Dr. Rick Russell in 2004 and has worked on studying the roles of a protein cofactor in group I intron RNA folding.

Permanent address: 200 Arlington Avenue. Apt BB Brooklyn, New York, 11208

This dissertation was typed by Amanda B Chadee